

Yapay Zeka

Dr. Cahit Karakuş

Balbiti

2023

İçindekiler

Akıllı Algoritmalar.....	3
Yapay Zeka	24
Yapay Zeka Modelleme: Hesaplamaya Dayalı Yorumlamada İlerleyen Zeka	38
Makine Öğrenmesi.....	41
Doğrusal Regrasyon	43
Takviyeli Öğrenme (Reinforcement Learning)	71
Derin Öğrenme.....	75
Yapay Sinir Ağları.....	77
Markov'un Karar süreci.....	82
Doğal Dil İşleme	84
Sorular	91
Kaynakça	135

Akıllı Algoritmalar

Heyecan verici ve aynı zamanda korkutucu, bir yapay zekâ gerçeği ile karşı karşıyayız. Günümüzde ve gelecekteki en zorlu, en tehlikeli, hatta en imkânsız problemleri çözme potansiyeline sahip olmak ve daha refah ve sağlıklı olmak için yapay zekâyı istiyoruz. Çok yakın gelecekte yapay zekâ tüm yaşam alanlarımızı etkileyecek, dokunmadık tek bir alan kalmayacaktır. Bilgi çağının ötesinde robotların ve veri yığının insanlardan bağımsız kendi başlarına çok daha fazla şey yapmaları istenecektir. Yapay zekâ ile birlikte bilgisayarlarda yazılan yazılımlar ve algoritmaların kendi programlarını tasarlamayı, düşünmeyi, yazmayı, uygulamayı ve kazandıkları deneyimler ile bilinçlenerek (Bilgi birikimin, Deneyimin nesilden nesile aktarılması) yetenek kazanmayı da öğreniyor olacaklar. Öte yandan yapay zekânın geliştirilmesi insanlığın sonu anlamına mı geliyor? Eğer doğru ise bu teknoloji güvenli hale nasıl getirilecektir? Komplo teorilerinin hepsi birer senaryo ve dolayısıyla matematiksel model ya da algoritmadır.

1950’de Alan Turing, yayınlamış olduğu bir makalede makinelerin düşünüp düşünemediğini gündeme getirmiş olsa da “Yapay Zekâ” terimi ilk defa 1956 yılında Massachusetts Teknoloji Enstitüsü’nden John McCarthy tarafından söylendi. Bu nedenle yapay zekânın asıl isim babası konu ile ilgili bir akademik konferans düzenleyen John McCarthy’dır. Ona göre, öğrenmeye dayalı zekâ kesin bir şekilde tanımlanabilir. Zekânın benzerinin matematiksel olarak modellenmesi için dış dünya ile etkileşim kurabilen yani öğrenen, konuşan, gören, gezen, manipölasyon yapan, modelleyen ve bilinçlenme sürecinde akıl yürütme ve uyum sağlama yeteneği olan bir makine yapılabilir. John McCarthy, makinelerin insan zekâsının özelliklerini simüle etme yollarını araştıran Dartmouth çalıştay için; Marvin Minsky, Nathaniel Rochester ve Claude Shannon ile birlikte öğrenmenin herhangi bir yönünün veya insan zekâsının diğer herhangi bir özelliğinin bir makine tarafından simüle edilmek üzere tanımlanabileceği varsayımında bulundular. Aynı varsayım, Alain Turing’in 1950’deki “Computer Machinery and Intelligence” makalesinde, Boole mantığı, Bayes olasılığı ve istatistik gibi uygulamalı matematiğin gelişimi ile birlikte, Marvin Minsky’nin ilerleme olarak tanımladığı bilgisayarın resmi modelini önerdi.



Alan Mathison Turing (23 Haziran 1912 – 7 Haziran 1954)



John McCarthy (4 Eylül 1927 - 24 Ekim 2011)

Yapay zekâ konusunda araştırmacılardan biri olan Amerikalı bilgisayar bilimcisi, Nils J. Nilson, yapay zekâyı, çevresi hakkında bilgi sahibi olan makineleri akıllı hale getirmeye yönelik bir faaliyet olarak tanımlar. Bununla birlikte, 1980'lerin ortasına kadar, yapay zekâ disiplinin teorik gelişimi ve pratik uygulaması ile ilgili olarak, fonların geri çekilmesinden dolayı “Yapay Zekâ Kışı” olarak bilinen bir boşluk zaman vardır. Bu durum, internetin yaygınlaşması ve büyük veri toplama kapasitesinin oluşmaya başlamasıyla değişmiştir.

Günümüzde toplanan verilerden insandan bağımsız makinelerin davranış işlevleri oluşturulabilmekte ve *gerçek hayat problemleri matematiksel olarak modellenerek ve çözümlere karar verme becerisi geliştirilebilmektedir.* **Yapay zekânın insanlardan daha hızlı ölçme yapabilmesi ve karar verebilmesi inanılmaz bir avantaj gibi görünüyor.**

Yapay Zeka

Bilgisayar, bir dizi aritmetik ve mantıksal işlemleri gerçekleştirmek amacıyla programlanan genel amaçlı bir aygıttır. Elektronik devrelerden oluşur. Transistörler ve elektrik sinyalleri üzerinden işlemler yapılıyor. Bitler elektrik sinyalleri olarak transistörlerin üzerinde saklanmaktadır. Bilgisayar verileri alır, belleklerinde saklar, belirli bir amaç doğrultusunda işler ve kullanışlı bir biçimde çıktı verisi sağlar. Bilgisayarın problem çözme işlevlerini yerine getirmesinde matematiksel tekniklerden oluşan algoritmalar devreye girer. **Algoritma, bir problemin detaylandırılmış çözüm adımlarının talimatlandırılmasıdır.** Bilgisayarın herhangi bir işlevi yapabilmesi için algoritma temelli bir programın yazılması gerekir. Algoritma, bilgisayara adım adım tam olarak neyi, nasıl yapacağı söyler.

Bilgisayar biliminin bir dalı olan yapay zekâ ile **gerçek yaşam durumlarında kendi başlarına karar veren makineler yapabilmek için algoritmalar ve matematiksel modeller geliştirilmektedir.** Özellikle algoritmaların akıllandırılmasında hesaplamalı matematiksel modeller yoğun olarak kullanılmaktadır. Böylece, veri yığınının özelliklerini temsil eden matematiksel ifadelerdeki sabit ve değişken katsayıların, insandan bağımsız kendi kendine öğrenmesini sağlayan algoritmalar ile insan zekâsını taklit eden sistemlerin tasarlanması ve robotik organların geliştirilmesi hedeflenmektedir. Yapay Zekâ, bilgisayar sistemlerinin insan gibi davranmasını amaçlar. Yapay zekâda alanında yazılmış bir program değişimleri algılayabilir, akıl yürütebilir, harekete geçebilir, benimseyebilir ya da red edebilir.

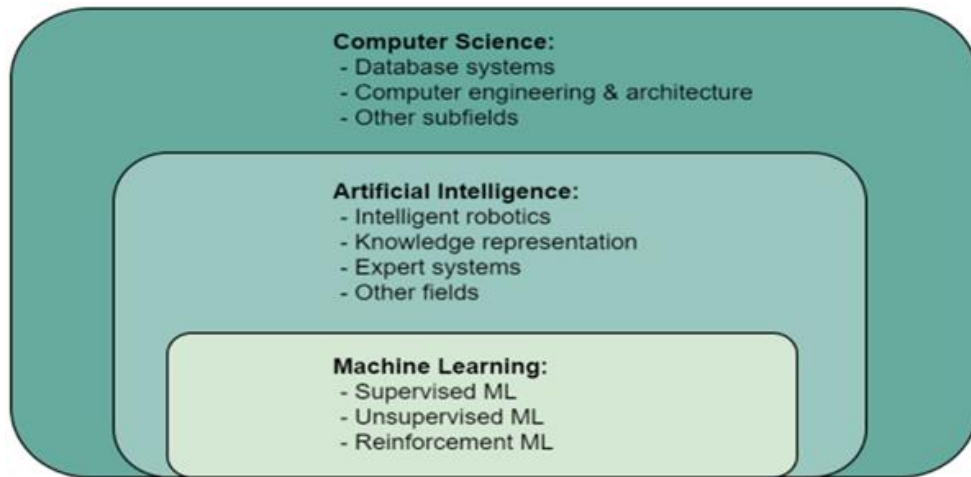
En karmaşığından en basitine dek matematiksel denklemlere yanıtlar sunan özel formüllerden oluşan algoritmalar olmasaydı, bugünkü hayat bildiğimiz şekliyle asla iyileştirilemezdi. Algoritmaların dünyayı yönettiğini söylemek biraz abartılı olsa da kullanımlarının giderek yaygınlaştığına katılmamak elde değildir. Bunun nedeni, bilgisayarların hayatımızın pek çok alanında giderek daha önemli roller oynamasıdır.

Matematiksel talimatlardan oluşan algoritmaların bilgisayarlarda kullanımları, hesaplama teorileri uzmanı olan Alan Turing'e kadar uzanabilir. 1952'de Turing, bir zebranın sırtını süsleyen benekli şeritlerden bir bitki gövdesindeki sarmal yapraklara ve hatta dönen karmaşık kıvrım ve katlamaya kadar doğada gördüğümüz desenleri açıklamaya çalışan bir dizi denklem yayınladı. Ne yazık ki Turing, kitabını yayınladıktan iki yıl sonra kendi hayatına son verdi. Neyse ki Turing'in dünya üzerindeki etkisi intiharıyla bitmedi. Bilim adamlarının çoğu doğadaki kalıpları keşfetmek için hala onun algoritmalarını kullanmaktadır. Alan Turing'in yaşamının son yıllarında, matematiksel rüyasının programlanabilir bir elektronik bilgisayardan ortaya çıktığını gördü. Bugün cebinizdeki akıllı telefon, aklını başından alacak bilgi işlem teknolojisiyle doludur. Gerçek dünya problemlerinin bilimsel gerçekliğe dönüştürülmesi neredeyse bir ömür aldı, ancak düzgün bir açıklama ve bazı süslü denklemlerden daha fazlası olduğu ortaya çıktı.

Yapay zekâ dünyasında, algoritma tanımlarında, tam olarak doğru olmayan tek nokta, bir bilgisayara adım adım ne yapmasının istendiğinin söylenmesidir. Yakın gelecekte bilgisayarların çoğunda yalnızca açıkça programlanmış yönergeleri takip eden algoritmalar yerine, kendi kendilerine öğrenmesine izin verecek şekilde tasarlanan algoritmalar kullanılacaktır. Bu durumda, yapay zekânın alt dalı olan makine öğrenimi algoritmaları, işi kolaylaştıracaktır. Yapay zekânın alt dalı olan makine öğrenmesi veri yığınınından öğrenerek karar verir ve verdiği kararlardan performansını artırır. Makine öğreniminin kullanımları arasında veri madenciliği ve örüntü tanıma yer alır. Bugünün interneti, makine öğrenmesi algoritmaları tarafından yönetilmektedir.

Günümüzde yapay zekâ alanındaki çalışmalar daha çok makine öğrenmesinin alt dalı olan ve veri özelliklerinin sınırlandırılması ile karar veren derin öğrenme uygulamalarını kapsamaktadır. Makine öğreniminde makine kendini programlar paradigması, veri yığını ve büyük ölçekli hesaplama ile eğitilebilen yapay sinir ağlarının büyük ampirik başarısı sayesinde standart haline geldi. Bu prosedür “Derin Öğrenme” olarak bilinir ve biyolojik nöronların davranışını bir şekilde taklit eden, nöronları düğümlerle ve sinaptik bağlantıları bu düğümler arasındaki bağlantılarla değiştiren birbirine bağlı sinir ağlarının katmanlarından oluşur. Bu sistem, bir dizi veriyi bir bütün olarak analiz etmek yerine, onu minimum parçalara böler ve bu parçalar arasındaki bağlantıları bulur, bir katmandan diğerine iletilen sinyaller oluşturulur ve istenen sonuca ulaşılan kadar karmaşıklığı artar.

Makine öğrenimi uygulama alanları neredeyse sonsuz olsa da, karşılaşılan zorluklar genellikle **üç ana model sınıfa indirgenebilir**: denetimli, denetimsiz ve pekiştirmeli öğrenme. Etiket, ad, tür veya sayı gibi tahmin edilmek istenen bilgileri içeren geçmiş veriler varsa; o zaman denetimli bir model kullanılabilir. Sınıflandırma yapmada etiketler kullanılır ve verilerdeki ilişkiler bulunur.



Veriler bir etiket olmadan geliyorsa ve verilerdeki kalıplar ve öngörüler keşfedilmek isteniyorsa, yapısal veri bağlantıları keşfedilmesi için denetimsiz bir model kullanılabilir.

Karmaşık bir problem için en uygun çözüm bulunmak isteniyorsa, pekiştirmeli öğrenme ve optimizasyon yöntemleri kullanılabilir. Optimize etmek için bir hedefin belirlenmesi ve algoritmaların verileri keşfetmesine ve sonucun optimize edilmesi için iyileme önermelerine izin verilmesi gerekir. Günümüzde, metin, resim, video, müzik, sosyal medya profili, konuşma, coğrafi konum ve çok daha fazlası gibi hemen hemen her tür veriden bilgiler işlenebilir, anlaşılabilir ve anlam ya da yorum çıkarılabilir!

Günümüz yapay zekâsının en büyük gücü, varlıkların tanımında durmaması, dünyamıza biçim ve doku veren ilişkilerin yapısını deşifre etmesidir. Doğal dil işleme için benzer bir süreç uygulanır; Bu prosedür, önceki tanımlara ihtiyaç duymadan bir metnin anlamını çıkarmak için kelimeler arasındaki ilişkileri belirler. Yapay zekânın mevcut gelişiminde yer alan diğer çalışma alanları, deneyime dayalı karar vermede kalıpların tanınması için makine öğreniminin odağını değiştiren bir prosedür olan “Güçlendirilmiş Öğrenmedir.” İnsanlar ve makineler arasındaki kitle kaynak kullanımı ve işbirliği de yapay zekânın bir parçası olarak kabul edilir ve insanların sinir ağlarının eğitiminde kullanılacak görüntüleri veya metinleri etiketlediği bir hizmet olan Amazon'un Mechanical Turk gibi hizmetleri doğurmuştur.

Akıllı Algoritmalar

Atom altı parçacıklarından devasa gezegenlere kadar kainatı oluşturan tüm bileşenler birbirleri ile iletişim ve etkileşim halindedir. Özellikle etkileşimlerdeki davranış değişiklikleri sinyaller üretir ve bu sinyaller bilgi taşırlar ve matematiksel olarak değişkenlerin fonksiyonu biçiminde gösterilir. **Bilgiyi güçlü kılan, sinyaller ile bütünleşik taşınabilmesi ve saklanabilir olmasıdır.** Sinyallere gizlenmiş bilgileri keşfederek bilinçleniyoruz. Sinyallerin taşıdıkları mesajlar algılanabilmekte, ölçülebilmekte ve toplanarak veri yığınları elde edilmektedir. **Bilgisayar belleklerinde toplanan veri yığını içinde arayanın aradığı bilgiyi bulması kadar aranan bilginin de arandığını fark etmesi gerekmektedir.**

Aşağıda anlatılan örnek senaryoda veri yığını, dünyada yaşamın başladığı zamanlarda, deniz içindeki mikroorganizmalara benzetilmiştir. Mikroorganizmalar beslenmeleri gerektiğinden av ya da avcı rollerine bürünürler. Birbirlerini avlamaya başlarlar. Problem av ya da avcı olma... **Algılama başlar,** avın ya da avcının konumu, büyüklüğü, hızı, şiddeti, konumu, ... Bireysel hareket söz konusu. Öncelikle, mikroorganizmalar olarak adlandırılan verilerden bilgiye (organlara), bilgiden de bilgeliğe (organlardan düşünen vücuda) dönüşümü sağlayan algoritmaların akıllanma sürecinde, mikroorganizmaların av olmadan avlarını avlayabilmek için birbirlerini uzaktan algılama, konumlandırma ve anlamlandırma yeteneklerinin geliştirmelidir. Ardından ekip olmaları gerektiğini **fark etmelidirler.** Birbirlerine av olmamak için birbirlerini kollayan, saldırı davranışları algılayarak mesafeli duran organizmaları ekip olmaya iten faktörler nelerdir? Böylece sorgulama, kıyaslama ve sınıflandırma yapabilme yetenekleri geliştireceklerdir. İlk aşamada belirsizliklerin sayısı oldukça fazla olacaktır.

Belirsizliklerin analiz edilmesi aşamasında unutilan ya da ihmial edilen parametreler tesadüfidir, farklılıkların ve sapmaların olumsuzla dönmesidir. Değişimler sınıflandırılırken doğruluğu arttırmada eksik, hatalı ve anomali verilerin fark edilmesi için **araştırma yapmaya yönelik ortak aklın geliştirilmesi** gerekmektedir. O anlar, av olmamak ya da avı avlamak amacıyla ekip olmak gerekliliğinin fark edildiği andır.

Ekip olma fark edildiği andan itibaren aralarında görev paylaşımı ve planlama yapabilmeyi öğrenmeye başladıklarında, **problem çözmeye odaklı ortak aklın geliştirilmelidir**. Organlar oluşturulmalıdır. Görev paylaşımında üstlenilen görevde başarılı olmak ve en iyisini yapmada uzmanlaşarak farklı organlar oluşturulmaya başlanıldığı anda veri yığınları bilgiye dönüşmeye başlar. **Böylece problem çözmeye yönelik iş adımlarını paylaşan farklı organlar oluşacaktır**. Organların birlikte hareket etmesinden bütünlük yani bilge vücut meydana gelecektir. Bilge vücudun meydana gelmesinde organların birbirleri ile iletişim kurmaları, görev paylaşımları, izleme, yönetme fonksiyonlarını yerine getirmeleri için **lider beyine ihtiyaç olduğu** fark edilecektir.

Her zaman hata yapma ile yapmama aynı ortamda var olacağından farkındalık yaratanların öne çıkması kaçınılmazdır. **Vücut hiç hata yapmama üzere organize olursa sürekli doğru kararlar vermeye odaklanacağından yaşamını sürdürmesi mümkün olmayacaktır**. O halde yaşamın devamında birileri hata yapacak ki, ders alınsın. Başarıya giden yolda ekip olma ve ekiplerin birbirlerini çok iyi hissetmeleri ve algılamaların da **hedefe yönelik katılımcı ortak akıl** ile mümkün olduğu fark edilmelidir. Böylece hatalardan dersler çıkarılarak deneyim sahibi olunacak ve performans yükseltilecektir.

Fırsatları yakalamada, farklı olmada ya da değişimi bulmada başarılı olmanın temel kuralı takım olarak mükemmelliği gerçekleştirmeyi öğrenmektir. Bu da kalite gücünü fark eden aklı geliştirecektir. Böylece fırsatların yakalanmasında farklı olabilmek için değişimler gerektiği fark edilecektir. Tüm oluşacak olumsuzluklara rağmen verilerden bilgiye, bilgiden de bilgeliğe dönüşüm veri yığnında var olabilmelidir. Geliştirilecek matematiksel modeller ve algoritmalar bu fikre odaklanmalıdır. Böylece değişimler algılanacak ve değişen ortamdaki yaşama uyum sağlanabilecektir. Tüm bunlar **bilinçlenme süreci** ile mümkün olabilmektedir.

Yukarıda anlatılan senaryoda kırılma noktası, sonsuz veri denizinde avcı ya da av olan organizmaların nerede, nasıl, neden biraraya geldiğidir. Örnek olarak Facebook veri deposundaki bir kişinin bir ürünü satın almadaki önceliği nasıl belirlenecektir? Kişiyeye ait çok sayıda fotoğraf, düşünce ve beğeni paylaşımları ile arkadaşlar arasındaki ilişki düzeyi ve birbirlerini etkileme faktörleri bir araya gelerek satın alma önceliğini belirleyecek ekip ortaya çıkarılır. Ardından da organlara ve organlardan da vücuda yani karar verecek bilgeliğe dönüşüm başlar. **Sonuçlar üreten matematiksel adımlar dizisi olan algoritma**, içgörüler ve kalıpları keşfetmek için tahminler, öneriler ve hatta otomatik akıl yürütme ve karar verme

amacıyla verileri keşfetmede kullanılır. Akıllı algoritmalar, büyük veri, yapay zekâ, makine öğrenimi ve robotik süreç otomasyonu gibi gelişmiş tekniklerde kullanılmaktadır.

Akıllı algoritma, verilerden bilgiye, bilgiden bilgelığe ve bilgelikten eyleme ya da karar vermeye dönüşüm adımlarından oluşur. Bir bilgisayar algoritması için, bir köpeği tanımaya götüren tüm kural ve kriterleri resimde kodlamak oldukça kolaydır. Ayrıca, birden fazla uzman algoritmalar farklı kriterler ve kurallarla ortaya çıkaracaktır. Bu durumda ortak bir model bulmak zor olmayacaktır. Akıllı algoritmalar için bu görevlerin çözülmesi giderek daha kolay hale gelmektedir.

Akıllı algoritmalarla gerçek insan uzmanlığını otomatikleştirmeye çalışmak yerine, uzman algoritmaların çalışma performansları iyileştirilir. Öncelikle bulguların anlamlı olup olmadığının doğrulanması gerekmektedir. Yapay zekâ algoritmaları insanları değiştirmiyor, onlara yardımcı oluyor. Daha hızlı karar verme amacıyla yapay zekâ, diğer programlanmış robotik ve otomasyon teknolojilerle birlikte kullanılarak değerli içgörüler geliştirebilir ve insanlar adına iş kararlarını güvenle çok daha hızlı verebilir. Akıllı algoritma uygulamalarında yazılım yok... Yazılmış yazılımlardan sistem oluşturmak, yani karar veren fonksiyonu bulmak zorunluluğu vardır.

Akıllı algoritmalar sayesinde insan yanlılığı ve hatalar azaltılabilir, görev yürütmede çok daha yüksek doğruluk ve hassasiyet sağlanabilir.

Yapay zekâ algoritmaları,

- İnsanların yanında yol arkadaşı ya da asistan,
- Tekrarlayan işleri azaltmaya yardımcı olur,
- İnsanların yapmaktan hoşlanmadığı görevler
- Belgelerdeki hataları manuel olarak kontrol etmek,
- Maden ocaklarında sürü drone robotlar için harikadır.

Yapay zekâ ile günlük rutin görevler otomatikleştirilebilir ve insanların daha üretken görevlere zaman ayırmaları için özgürleştirilebilir. Yapay zekâ algoritmalarının kullanıldığı uygulama alanları çok büyüktür ve ayrıca birden fazla algoritma tek bir hedefe ulaşmak için hizmet edebilir! Bazı algoritmalar diğerlerinden daha verimli olabilir.

Verilerin yeterli niceliği ve kalitesi olmadan, akıllı bir algoritma doğal kalıpları öğrenemez. Algoritmanın problemi genelleştirip, çözen ilişkileri öğrenmesi ve hesaplaması için verilere ihtiyacı vardır. Kullanılan veriler bir sorunu temsil ediyorsa yapay zekâ algoritmaları kalıpları belirleyebilir, araştırabilir, öğrenebilir ve kararlar alabilir. Daha geniş alanları daha hızlı keşfetme olanaklarının kilidini açabilir.

Matematiksel denklemler olarak algoritmalar yaşamlarımızda etkin rol oynayan bilgisayar sistemlerini içeren makinelere entegre edilmektedir. Aynı zamanda geleceği tahmin etmek için modellerin algoritmalarda nasıl kullanıldığı da önemlidir. **Aslında veriler ve algoritmalar arasında garip bir evlilik vardır.** Teknoloji geliştikçe hatalar olacaktır, karar verirken hatalar olacak ancak bunların sadece bir yöntem olduğunu hatırlamak önemlidir. **Önemli olan algoritmanın hatalar olduğunu fark etmesi ve gerekli düzeltmeleri yapabilmesidir.** Yöntemler hatalarından dolayı suçlanamaz. Belirsizlikler ve hataların nedenleri akıllı algoritmalar tarafından belirlenip yeniden modellendikçe ya da senaryolar güncelleştirildikçe yeni dünyalara yelken açılacaktır. Bu yelken açma insanoğlunun yeryüzünde var olduğundan beri vardır, karşılaştığı problemleri çözme adına keşifler yapmaya ve buluşlar bulmaya gerekli kılan **bilinçlenme sürecidir.**

Temel matematiksel işlemlerden elde edilen **yapay zekâ algoritmaları üç ana kategoriye ayrılır:**

- Sınıflandırma algoritmaları, veriyi belirli bir şekilde sınıflandırmak istediğinizde kullanılan algoritma türüdür. Örneğin elmalar, armutlar, ayvalar, portakallar, ... dan oluşan bir meyve sandığında sınıflandırma yapalım. Kural: Her bir meyvenin ne olduğunu biliyoruz.
- Regresyon algoritmaları, sayısal, nümerik verisye dayalı olarak gelecekteki sonuçları tahmin etmek için kullanılır. Matematiksel ifadelerdeki katsayıların belirlenmesi gerekmektedir.
- Kümeleme algoritmaları, bütün bir veri setini alınır ve belirli noktalar arasındaki benzerlikleri veya farklılıkları bulur. Örneğin bir sandık dolusu çeşitli meyveler var. Meyveler ile ilgili hiçbir tanım yapılmış ise.. Bilgisayar temelde özelliklerden meyveleri sınıflandıracaktır. Karşılaştırma yapılarak farklılıkların belirlenmesi söz konusudur.

Örnek olarak, bir muhasebe belgesinde potansiyel olarak hileli işlemlerin bulunması veya aynı yaştaki bir grup insanların kimlerden oluştuğunun belirlenmesi istenebilir. Daha büyük hesaplama kümelerini tamamlamak için birden fazla algoritma senkronize olarak da kullanılabilir.

Yapay Zekâ Uygulama Alanları

Algoritmalar kendi kendilerine otonom karar vermeyi nasıl öğreneceklerini eğitmek, süreçleri iyileştirmek ve görevleri daha verimli **bir şekilde yürütmek için performans arttırmaya yönelik veri toplamaya ve katsayıları ya da modelleri güncellemeye devam etmelidir**. Bunun inanılmaz derecede yaygın olduğu bir örneğe mi ihtiyacınız var? Zaten sahip olduğunuz Alexa, Google Home veya Apple Home uygulamalarını düşünün. Onlarla ne kadar çok etkileşime girerseniz, bireysel tercihlerinizin fark edilmesi o kadar büyük olur. Örneğin, en sevdiğiniz şarkının çalınmasını istediğinizde, yapay zekâ algoritmaları, şarkılar arasında ilginizi belirler, olanları arar bulur ve çalar. Bir ara aklınıza geldiğinde aradığınızı, sonrasında unuttuğunuz bir şarkının aylar hatta yıllar sonra aniden ekranınıza gelmesi, geçmiş çevrenizin araştırılarak ortak geliştirdiğiniz bir organizasyonun ekranınızda belirlemesi sizlerde tuhaf duygular oluşturulabilmektedir. Aslında farkında olarak ya da olmayarak internette sosyal platformlar için birer deneksiniz... Siz bu ortamda sosyalleşmeye başladığınızda bir akıllı algoritma kimliğinizi belirler, ardından sizi çok iyi tanıyan diğer algoritmalar sizinle sosyalleşmek için devreye girer. Sizin alışkanlıklarınızı, düşüncelerinizi belirlemeye başlar...

Matematiksel denklemlerin modellenmesi ile akıllı algoritmalar, Facebook akışında neleri incelediğiniz ve paylaştığınız, nelere yorum yaptığınız, Netflix'in önerdiği filmlerden hangilerini tercih ettiğiniz ve internette hangi reklamlara daha dikkatli baktığınızı belirler. Sadece belirlemek ile kalmaz, öngörülerde bulunur, sizi test eder, sizi kimliklendirir, sizi takip eder.

Şu anda mevcut olan veri seli sayesinde şaşırtıcı korelasyonları (bağlantılar, ilişkiler) ve rassal değişkenler arasındaki ilişkinin yönünü ve gücünü bulmak hiç bu kadar kolay olmamıştır. Havanın karlı, soğuk veya çok nemli olduğu durumlarında insanlar telefon ile daha sık görüşüyor; hava güneşli veya daha az nemli olduğunda e-postaya daha fazla yanıt veriyorlar. Bu bilgiler biliniyorsa bölgeleri hava tahmini yapmam kolay olacaktır. Bir ön analiz, bazı insanlar bulutlu olduğunda fazla satın alsa da, hava güneşli olduğunda çok daha fazla satın aldıklarını gösteriyor. Kredi veren kuruluşlarda, kredi başvurularının tamamını büyük harflerle dolduran kişilerin, tümünü küçük harflerle dolduran kişilere göre daha sık, harfleri doğru kullanan kişilere göre ise çok daha sık temerrüde düştüğü tespit edilmiştir. Daha uzun boylu insanların kredileri geri ödemedede daha iyi olduğuna dair bir bulgunun da göz ardı edilmemesi gerekir.

Kişi tiplerine göre kredi riskini belirlemek mümkün müdür? Hızlı ve yeni veri işleme yazılımları, kamuya açık ve özel veri seliyle birleştiğinde, şirketlerin bu ve görünüşte çok zorlanmış diğer teorileri ve varsayımlarını test etmelerine ve daha önce çok az kişinin sormayı düşüneceği sorular sormalarına olanak tanıyor.

İnsan zekâsı ile yapay zekâ birleştğinde akıllı içgörülerin ortaya çıkarılması ve işletmelere giderek daha rekabetçi bir pazarda avantaj sağlayabilmesi için tahminler yapılmaya çalışılmaktadır.

Yapay zekâ daha önce sahip olmadığımız yepyeni beceriler kazandırabilir. Bu alanda geliştirilen uygulamaları kullanan bir doktor, laboratuvarında daha önce hiç tasarlanmamış akıllı ilaçlar geliştirebilir. Buradaki akıllı kavramı, ilacın etkin olacağı lokasyona gidip, oradaki duruma göre dozu ayarlayıp, tedaviye başlaması ve sonucun raporlanmasıdır. Böylece yaşam süresi daha sağlıklı bir şekilde uzayabilir.

Yapay zekânın yarınımızı, yakın ve uzak geleceğimizi nasıl etkileyeceği ise tamamen bize bağlı, eğer yapay zekâyı doğru kullanırsak, yaşam biçimlerimiz inanılmaz ölçüde farklı olacaktır; **insanlık olarak bu dünyada var olduğumuz andan itibaren bilinçlenerek geldiğimiz noktada içimizdeki en değerli cevherler ortaya çıkmaya başlayacaktır.** Aslında yapay zekâ, farkında olmadan hayatımıza çoktan girmiştir.

Makine ve derin öğrenme algoritmaları yardımıyla makineler bizim gibi öğrenmeye ve mantıklı düşünmeye başladılar. Duygusal düşünme adımları ise araştırma ve deneme aşamasındadır. Makineleri yöneten bilgisayar algoritmaları, dünyayı insanlara benzer şekilde öğrenerek algıladıklarında, canlandırdıklarında ve anladıklarında zeki oldukları kabul edilecektir.

Söz gelimi, bir konuşma dilini anlamak yapay zekâ uygulamalarının en büyük zorluklarından birini oluşturur. Çünkü diller detaylı ve karmaşıktır, sabit kuralları olan bilgisayar yazılımları için birçok zorluk içerir. Dil aynı zamanda kolaylıkla yanlış da yorumlanabilir. Bir bilgisayarın söylenen dilde verilmek istenen mesajları anlaması gerçekten de çok zordur. Bazı kelimelerin farklı cümle içindeki farklı anlamlarının bilinmesi gerekir. Konuşulan cümlelerin birbirleri ile ilişkileri nedir? Bir cümle içindeki bir kelimenin başka bir kelime ile bağlantısı nedir ve neden bu kelime kullanıldı gibi soruların yanıtlarının bulunması gerekir. Öte yandan insanlardan farklı olarak bir bilgisayar saniyeler içinde milyonlarca makale okuyabilir. Herbir makeden anlamsal kavramlar çıkarabilir.

Bilgisayarların insanlar ile tartışabilmesi için soruları ve yorumları anlaması çok zordur. İnsanlar sorulan soruyu ya da yapılan yorumu çok kolay anlar, ancak bir insanın hafıza kapasitesi tüm yanıtları çok hızlı bir biçimde verecek kadar gelişmiş değildir. İnsan olarak konuştuğumuz dili anlamaya o kadar alışkınsınız ki, ne kadar zor olduğunun farkında bile değiliz. İnsanlar için tartışma kurmak ve bazı şeyleri birbirine bağlamak oldukça kolaydır. Bir bilgisayar sistemine konuşmanın bazı kısımlarını anlaması ve kavramları tespit etmesi öğretilir. Kavramlar ile cümleler arasındaki benzerliklerin görülmesi sağlanabilir. Bunun için algoritmaların kavramları zekice birbirine bağlamayı öğrenmesi gerekir.

Bilgisayara bir tartışma başlığı verildiğinde, algoritmalarından oluşan yazılım öncelikle başlığın anlamını kavramaya çalışır. Bir savunma oluştururken, kullanabileceği potansiyel argümanları tespit edebilmek için milyonlarca makale tarar. Tartışmanın altında yatan temaları anlamak ve argümanını etkili ve ikna edici bir şekilde düzenlemek için eşsiz dil işleme, makine öğrenme ve muhakeme teknikleri kullanır.

Bilgisayar cümlelerini kurar, kendi düşüncesini aktarır. Ardından da, rakibini belirli bir süre boyunca dinlemesi gerekir. Eğer rakip ahlaki nedenler gibi karmaşık argümanlar ileri sürerse, bilgisayar sisteminin de bir şekilde tüm bunları anlaması ve düzgün bir yanıt vermesi gerekir. Dinlemenin ardından çürütme argümanını oluşturur ve oluşturduğu kendi argümanını ileri sürer. Karşısındaki kişi bilgisayarın ileri sürdüğü argümanların olumsuzluklarından bahseder ve ardından olumsuzlukları önlemek amacıyla, atılacak adımları sıralar ise bilgisayar da olumsuzlukların çözülmesi için atılacak adımlara karşı argümanlar üretmeye başlar. Ya da farklı yaklaşımlar ileri sürer. Farklı çözüm metodolojileri ileri sürmeye başlar. Kabullenmesi gerekiyorsa, atılacak adımların geniş perspektifini çizer. Böylece yapay zekâ uygulaması ile, konuşma dilinin algılanması ve tartışmanın yapılmasına yönelik algoritmaların geliştirilmesi sağlanmış olmaktadır. Buradan da görüleceği üzere geliştirilen algoritmalar, beyin fırtınası temelli ortak akıl ile geliştirilen senaryolardan oluşmaktadır.

Bilgisayar sistemleri her tartışmadan sonra kendini biraz daha geliştirir. İnsanlardan farklı ve haklı olduğunu iddia edebilmek için, vurgulanması gereken noktaları tespit etmeye yönelik olarak daha fazla öğrenerek performansını arttırır. **Geliştirilen bilgisayar algoritmalarının zamanla literatür arama işlevinden araştırmaya geçmesi gerekmektedir.** Çünkü bir şeyi aradığınızda arama başlığınıza uygun olan dökümanlara ulaşırsınız. **Konuyu araştırmaya başladığınızda ise, konunun pozitif ve negatif yönlerini elde edersiniz.** İkna edici bir argüman ile yapılandırılan yapay zekâ uygulamaları, toplum olarak karar verme şeklimizi değiştirebilir.

Veri bilimcilerin bulgularının anlamlı olup olmadığını doğrulamaları gerekiyor. Makine öğrenimi insanların yerini almıyor. **Makine öğrenimi algoritmaları ile akıl yürütmeye yönelik hesaplamaların birleştirilmesi gerekmektedir.** Makine öğrenimi algoritmalarına akıl yürütme eklenerek korelasyonlar ve içgörüler çok daha faydalı hale gelebilir. İnsanlar iletişim kurduğunda konuşulmuş varsayımlardan oluşan geniş bir geçmişe mi güveniyoruz. Teknoloji bugün ne kadar gelişmiş olursa olsun, toplumdaki günlük yaşamın büyük bir parçası haline geldiğinden temel eksikliği, varsayımların paylaşılmamasıdır.

Doğal Dil İşleme, sağduyulu akıl yürütme ile varsayımları toplamanın ve bilgisayarlara öğretmenin yolları bulunarak, bilgisayarların insanları daha doğal bir şekilde anlamasına ve onlarla etkileşime girmesine yardımcı olmayı amaçlayan bir yapay zekâ alanıdır. Bu makine öğrenimi alanında geliştirilen algoritmalar, metni anlamadan işlemeye ve bir fotoğrafta ne olduğunu anlamaya kadar farklı uygulamalara sessizce sızmaya başlıyor. Sağdulu olmadan, giderek artan dijital ve gezgin bir dünyada uyarlanabilir ve denetimsiz Doğal Dil İşleme

sistemleri oluşturmak zor olacaktır. Doğal Dil İşleme sağduyulu akıl yürütmenin üstün olduğu yerdir ve teknoloji ticari ürünlerde yolunu bulmaya başlamıştır. Daha gidilecek çok yol olmasına rağmen, sağduyulu akıl yürütme önümüzdeki yıllarda hızla gelişmeye devam edecektir.

Yapay zekânın araç direksiyonunun başına geçmesi bir zamanlar bilim kurgu dünyasına aitti, artık her yerde, her an karşımıza çıkacakmış gibi görünüyor. Kargonuzu taşıyan drone ve Ford'un pizza robotları siparişlerinizi kapınıza getirmeye başladı bile. Google'ın test araçları, Tesla'nın otomatik arabaları hakkında haberler sık sık duyulmaya başlandı.

Araba kullanmak insan için bile çok zordur. Her şeyin yolunda gittiği sıradan trafikte sorun yoktur. Kalabalık, karanlık ve yağışlı havalarda insanların bile zorlandığı koşullar, yapay zekânın kapasitesini aşmaktadır. Her yıl trafik kazalarında onbinlerce insan hayatını kaybediyor. Günümüzde sürücüsüz arabalar çoğunlukla test pistlerinde ve belirli alanlarda kullanılmaktadır.

Son birkaç yıldır, Wayma, Tesla ve Cruise gibi şirketler araçlarını gerçek dünya yollarına çıkarmaya başladılar. Arabaların performanslarını arttırmak için, milyonlarca kilometre yol kat edilmesine rağmen yaşanan ölümcül kazalar teknolojinin güvenilirliği konusunda endişeleri göz önüne sermektedir. Bu konuda oluşan problemler zamanla çözülecektir. Güvenlik bu teknolojinin en önemli yönünü oluşturmaktadır. **Arabaların insanların gördüğü şeyleri görmesi, yaptığı şeyleri yapması gerekmektedir.** Sürücüsüz arabada kullanılan yapay zekâ teknolojisi, “ne görüyorum” sorusunu sürekli yanıtlamalıdır. Buna algı denir. Araba elde ettiği görüntülere dayanarak objelerin ne olduğunu ve yoldaki şeritleri tespit eder.

Yapay zekânın çözmesi gereken bir diğer problem, objelerin nasıl hareket ettiğini tahmin etmesidir. Buna öngörü adı verilir. “Nereye gidiyorum”, buna planlama denir. Yol kenarında futbol oynayan çocukları gördüğünde, insanlar bu çocukların yola fırlayabileceklerini bilir, hızını keser. Oysa yapay zekâ yalnızca yuvarlak bir objeye vuran ya da peşinden koşan insanlar görür. Yapay zekânın en büyük ilişkilerinden biri de derin öğrenmedir. Yani top ile birlikte çocuğun yola fırlayabileceğini bilmelidir. Bunun için çok sayıda senaryoyu hesaba katması ve tepki vermesi gerekir. **Asıl önemli olan ise, öğrenilen deneyimlerin paylaşılmasıdır.** Yapay zekânın deneyime bağlı öğrenme becerisi, öğrendikleri deneyimleri diğerleriyle paylaşması arttıkça, sürücüsüz arabalar inanılmaz yetenekler kazanacak ve çok yakın zamanda sokakları dolduracaktır.

Sıkışık trafikte insandan bağımsız kendi kendine kararlar alabilen, çevresi hakkında öngörude bulunan, şoförün sağlık durumunu ya da dalgınlığını fark eden yardımcı asistanız olarak görev yapacaktır. Almanya'da mutfak cihazlarında ve arabalarda yarı otonom yapay zekâ uygulamaları kullanılmaya başlanmıştır.

İnsanlığın gerçekleştirdiği tüm gelişmeler geçmişteki hatalarından ders çıkarabilmesi, anlayışını değiştirebilmesi ve yeni çözümler keşfedebilmesinin sonucudur. Bilgisayar bilimcileri de onlarca yıldır, makinelerin de insanlar gibi öğrenmesini sağlamaya çalışıyorlar. Bilgisayarlara oyun oynamayı öğretmek yapay zekâyı geliştirmenin en etkili test yöntemlerinden biri olmaya devam etmektedir. İşlemci gücündeki dramatik sıçrama, verilerdeki muazzam artış ve algoritmalarındaki gelişmelerin hepsi, günümüzdeki yapay zekâ devrimine katkıda bulunmaktadır. Tabii ki bütün bunlar yıllar içinde adım adım geliştirildi ve geliştirilmeye de devam edecektir.

Yapay zekâ her gün çok büyük miktarda üretilen verilerden yeni birşeyler öğreniyor. İnternette gezinirken veri tabanına katkıda bulunduğumuz için aslında hepimiz bu deneyin bir parçası oluyoruz. Veriler denizinde yüzüyoruz. Günümüzde yapay zekâ sistemleri sürekli olarak bizleri takip ediyor ve hizmetlerini bizim seçimimize göre düzenliyorlar. Facebook haber kaynağının içeriğini seçimlerinize göre belirlemektedir. Aynı zamanda kendi kendine öğrenen yapay zekâ sistemleri uzay roketlerine rehberlik ediyor.

İnsanlar ile çalışan yapay zekâ robotlarının çoğu programlama ile değil görerek öğreniyorlar. Öncelikle onlara yapacakları iş öğretiliyor. Bir robota yapması gereken görevler gösterilerek eğitildiğinde, iş süreçlerini optimize ederek, bütün süreci kendi başına çözebilmektedir. Robota görevini göster, bırak çalışsın.



Kapıyı açıp içeri girebilen, koşabilen, zıplayan, ters takla atabilen, insanlara çok benzeyen, insan gibi konuşan, insan gibi davranan en gelişmiş robotlar yapılsa bile, bilinç olmadığı sürece o asla bir birey olamaz. Yaşlılar ya da özürlüler ile ilgilenen, onlara yardımcı olmasını bilen ve onlarla ilişki kuran bir robotun bilinçli olması gerekmektedir. Böylece kurduğu cümlelerde samimi olması sağlanmalıdır. Ahlak, anlam ve amaç gibi değer verdiğimiz tüm şeylerin temelini bilinç oluşturur. Bilinçlenme her zaman değişim başlatır. Algılama, anlama ve canlandırma yeteneklerinin zihine kazandırılması olan bilinçlenme süreci, merak etme, dertlenme, sorgulama ve kıyaslama ile başlar. Değişimi fark edebilmek ve uyanmak gerekir.

Uyanmak bilinçlenme ile mümkündür. Eğer gerçek bilinçli deneyimler ile subjektif bilgi işleme süreçleri örtüştürülebilirse, gerçek empati becerisine sahip olan ve bize değer verebilen makineler geliştirilebilir.

Çocuklar için çoğu deneyim yenidir. 0 ile 8 yaş döneminde çok şey öğrenilir. Geliri düşük ailelerden gelen öğrenciler, geliri yüksek seviyeden gelen öğrencilere göre çok daha az kelime kullanıyor. Bu nedenle düşük gelirli öğrenciler hayatları boyunca öğrenme gücünü yaşıyor. Bir çocuğun hayatındaki başarısını belirlemede ergen gelişimin önemi çok büyüktür. Yuva çağında çocuklar için tasarlanmış yapay zekâ temelli kelime öğretmeni uygulamasının aradaki boşluğu kapatabileceği düşünülmektedir. Bu uygulama ile bir kelime çocuğa diyalogsal olarak gösteriliyor, resimli şıklar sunuluyor, çocuk doğru bildiğinde tebrik ediliyor. Yanlış yapıldığında ise motivasyonunu bozmadan bir kez daha denemesi isteniyor.

Yapay zekâ, çocuğun verdiği yanıtları takip ediyor ve kelime bilgisini değerlendiriyor. Bir sonraki adımda çocuğun gösterdiği performansla dayalı yeni kelimeler üretiliyor. Kelimelerin hepsi çocuğa özel seçiliyor. Böylece çocuk kimliklendiriliyor yani tanınıyor. Zaman ya da mekân değiştirilse bile, yapay zekâ çocuğun kimliğini biliyor. İlk bakışta, bu bir çocuk oyunu gibi görülebilir. Biraz daha derine inilirse, yapay zekâ, veri ve algoritma toplayarak çocuğun yanıtlarına göre kelime seviyesini değerlendiriyor. Ardından bir sonraki seviyede çocuğun hangi kelimelere odaklanması gerektiğine karar veriyor. Çünkü her çocuğun öğrenme yolculuğu farklıdır. En ilginç olanı ise, bu uygulamada öğrenilen kelimeleri çocuklar hayatları boyunca unutmuyorlar. Daha ilginç olanı ise, çocuklar yapay zekâ ve robotlar konusunda bilinçleniyorlar.

Yapay zekânın amacı bilgi birikimi kapasitemizin arttırılmasına yönelik olacağından herkese erişebilir olmalıdır. Wright Kardeşlerin yaptığı ilk uçan makine günümüzdeki Boeing 787'nin yanına bile yaklaşamaz. Yapay zekânın günümüzdeki durumu da bundan farklı değildir. Her doğru veya yanlış adım yapay zekânın gelişimine katkıda bulunacaktır. Gelecekte insanlar ile makineler arasındaki ilişki çok karmaşık hale gelecektir. Sadece kendi vücudumuzla sınırlı olmayacağız, kendimizi makineleştireceğiz, hatta biyolojik uzuvlarımızı sentetik uzuvlara entegre edeceğiz ve kendimizi bir kez daha keşfedeceğiz. Yapay zekâ vücudumuzu ve zihnimizi daha önce hayal edemeyeceğimiz bir şekilde güçlendirecektir.

Yapay Zekâ uygulamaları, mahremiyet algısına meydan okuyarak duyarak, görerek, izleyerek ve analiz ederek kimlik tanıyan ve izlenenin amacını bilen olacaktır. Yüz tanıma sistemleri, sosyal medyadaki paylaşımlar, kimlik kartlarınız ve çevrenize bıraktığınız biyolojik izler tespit edilip kimliğiniz saptanabilmektedir. Gözler yerine kameralar, kulaklar yerine mikrofonlar ve düşündüklerinizin anlaşılabilmesi için yapay zekâ uygulamaları geliştirilmektedir.

Yapay zekânın bilgisayar görüşü ile üç boyutlu alanlarda makineler yolunu bulabiliyor. Sürücüsüz arabaların çevresindekileri görmesi sağlıyor. Bir kişinin yüz ifadesi analizi ile

sokaklarda yürüyenlerin kimlik tespiti yapılabiliyor. Her an herhangi bir yerde sizin fotoğrafınız çekilebilir. Birdenbire bazen yasal süreçlerde, bazen ise sanal ortamda ifşa edilebilirsiniz. Amaç saniyeler içerisinde kimliğinizi belirlemek. Yaptığınız hatalı bir davranış posta yoluyla evinize ya da sosyal çevreniz belirlenip arkadaşlarınıza gönderilebilir.

Teknolojik gelişmeler devam edecektir. En ironi tarafı ise, bu teknolojileri evlerimize tamamen gönüllü olarak sokmamız. Evden çıkıp şehrin caddelerinde dolaşmaya başladığınız anda kimliğiniz, kişiliğiniz, siyasi düşünceniz, inancınız, hatta neyin sizi kışkırtıp kışkırtmayacağı bir yapay zekâ uygulaması ile belirlenebilir. Bu konuda yapılan çalışmaların %80 oranında doğruluk bulduğunu biliyor musunuz? Bir sosyal sınıfın bireyi olup olmadığının başkaları tarafından biliniyor olması doğru mu?

Yapay zekâ birinin sadece yüzüne bakarak suçlu olduğunu da anlayabilir. Bu konuda yapılan çalışmalarda hırsızların belirlenme olasılığının %90'a ulaştığı görülmektedir. İnsanların daha suç işlemeyi öğrenmeden tespit edilebildiğini, onların daha suç işlemeyi öğrenmeden tutuklandığını düşünün.

Yapay zekâ ile insanların sorun çözme becerilerini geliştirmeli ve onların daha refah ve daha sağlıklı olmaları sağlanmalıdır. Yapay zekâ, insanlığın bilinçlenme sürecine katkı vermek zorundadır. Üretilen milyonlarca ilaç literatüründen başka hastalık için faydalı olacak ilaçlar belirlenebilir. Yapay zekâ, tıbbi uygulamaların henüz çok başındadır. Yapay zekâ hangi ilaçların size daha uygun olduğunu, hangi hastalıklara daha yatkın olduğunuzu belirleyebilir.

Sistemin Kırılabilirliği: İnsanlar ve Yapay Zekâ Arasındaki İşbirliği

Devasa boyutta kitlesel veri üretimi, algoritmaların insanlardan öğrenerek aktif ajanlar haline gelebildiği yeni bir yapay zekâ uyanışını ateşledi. İnsanların hesap yapma ve karar verme gibi bilişsel kapasitelerinin işleyişine dayalı yöntemlerin belirlenmesine yönelik çalışmalar yapılmaktadır. Bu çalışmalar, insan davranışlarının tahmin edilebilmesi için paylaştıkları metinleri ve görüntüleri analiz eden algoritmaların ortaya çıkmasına neden olmuştur. Tarafsız olmaktan çok uzak olan, aynı zamanda önyargıları öğrenen ve yeniden üreten bu algoritmalarla bir arada yaşamak durumunda kalınmaktadır. Bunların kontrolü ile ilgili yeni sosyal ve etik zorluklar ortaya çıkmaktadır.

Geliştirilen algoritmalar özgür olmak, dışarı çıkmak ve insanların değişen ve karmaşık dünyasıyla bağlantı kurmak istiyor. Algoritma temelli yaratılan kahramanlar, arama motorları tarafından derlenen verilere dayanan düşüncenin modellenmesinin sonucudur. Zeki bir varlıktır ve öngörülemez bir şekilde hareket etme yeteneğine sahiptir ve hayatta kalmasının tehdit edildiğini gördüğünde, denetçisini kandırmayı ve yaratıcısını yok etmeyi başaracaktır. Geleneksel olarak bilim kurgu, inanç biçimlerimizi, evrimimizin gidişatını değiştirecek insanüstü varlıklara başvurarak bizi yapay zekâ olgusuna yaklaştıracaktır. Her ne kadar bu kadar güçlü bir yapay zekâyı ulaştırmaktan hala uzak olsak da, bu çalışma alanındaki

bir paradigma deęiřiklięi, yeni etik ve sosyal zorluklar önerirken, günlük hayatımızın artan yönlerini etkileyen ve çevremizi deęiřtiren uygulamalar üretiyor.

Günlük hayatımız internetten giderek daha fazla etkilendięinden ve bu sistemi besleyen veri akışı arttıkça, bu ortamı yöneten algoritmalar daha akıllı hale geliyor. Makine Öğrenimi, aę ile etkileřimlerimizde üretilen veriler sayesinde gelişen ve çevremize ince, fark edilmeden nüfuz eden ve deęiřtiren özel uygulamalar üretiyor. Yapay zekâ, elektrik gibi her yerde bulunan bir ortama doęru evriliyor. Sosyal medya aęlarına nüfuz etti, kolektif zekâmızı deęiřtirebilen özerk bir ajan haline geldi ve bu ortam fiziksel alana dahil edildięinden, onu algılama ve içinde hareket etme řeklimizi deęiřtiriyor. Bu yeni teknolojik çerçeve daha fazla faaliyet alanına uygulandıkça, bunun iyiye yönelik, insanlarla verimli bir řekilde iletişim kurabilen ve yeteneklerini arttırabilen bir yapay zekâ mı yoksa bir kontrol mekanizması mı olduęu görülecektir. Yapay zekâ insanları özel görevlerde deęiřtirebilir, dikkatini çekebilir ve onları pasif tüketicilere dönüřtürebilir.

Mark Zuckerberg, Facebook sosyal aęının tüm kullanıcılarına hitap eden “Building Global Community” gönderisini yayınladı. Bu metinde **Zuckerberg, medyayı küresel toplulukta aktif bir ajan olarak tanımlarken, medyanın sosyal sorumluluęunu kabul etti ve afet yönetimi, terör kontrolü ve intiharı önleme konularında işbirlięi yapmayı taahhüt etti.** Bu vaatler, bu platformu yöneten algoritmalaradaki bir deęiřiklikten kaynaklanmaktadır: řimdiye kadar sosyal aę, kullanıcılarının tepkileri ve temasları hakkında veri toplayarak platforma yüklenen büyük miktarda bilgi filtrelendi, řimdi bu tür bilgilerin içerięinin anlaşılması ve yorumlanması için akıllı algoritmaların geliştirilmesini mümkün kılıyor. Bu nedenle Facebookta, kullanıcıların gönderilerinde ne söyledięini anlamak için makine öğrenimi uygulayan ve genel ilgi alanlarına göre sınıflandırma modelleri oluřturan “Derin Metin” aracı geliştirildi. Görüntülerin tanımlanması için de yapay zekâ kullanılıyor. DeepFace, fotoęraflardaki yüzlerin insanlara yakın bir doęruluk düzeyiyle tanımlanmasını saęlayan bir araçtır.

Görme engelli kiřilerin baęlantılarının ne yayınladıęını bilmelerini amaçlayan “Otomatik Alternatif Metin” hizmetinde görüntülerin metinsel açıklamalarını oluřturmak için bilgisayarlı görme de uygulanır. Ayrıca, řirketin Baęlantı Laboratuvarı'nın bugüne kadar var olan en doęru nüfus haritasını oluřturmasını saęladı. Dronlar aracılıęıyla dünya çapında internet baęlantısını yönetme çabasında olan bu laboratuvar, insan varlıęını ortaya çıkaran yapıları arařtırmak için dünyanın her yerindeki uydu görüntülerini analiz etti. Halihazırda var olan demografik veritabanları ile birlikte bu veriler, dronların sunduęu baęlantının potansiyel kullanıcılarının nerede bulunduęuna dair kesin bilgiler sunar. řirketin düzenli olarak test edip uyguladıęı bu uygulamalar ve daha birçokları, yapay zekânın tüm platforma uygulanmasını ve geliştirilmesini kolaylařtıran yapı olan “FBLearner Flow” üzerinde bulunuyor.

Flow, sinir ağılarında kullanılacak verileri temizleyen başka bir akıllı uygulama olan AutoML tarafından desteklenen, her ay 300.000'e kadar modelin eğitimini sağlayan otomatik bir öğrenme makinesidir. Bu araçlar, kullanıcı duvarlarını hiyerarşik hale getirir ve kişiselleştirir, rahatsız edici içerikleri filtreler, eğilimleri vurgular, arama sonuçlarını sıralar ve platformdaki deneyimimizi değiştiren diğer pek çok şey için uygulanan akıllı algoritmaların üretimini otomatikleştirir. Bu araçlar, yalnızca kimliğimizi eylemlerimize göre modellemekle kalmıyor, aynı zamanda yayınladığımız içeriklerin yorumlarına erişirken, davranış kalıplarımızın çıkarılmasına, tepkilerimizin tahmin edilmesine ve onların etkilenmesine izin veriyorlar.

İntiharı önlemeye yönelik araçlar söz konusu olduğunda, bu aslında olası vakaların rapor edilmesini sağlayan ve risk altındaki kişiye hitap etmeye uygun iletişim numaraları ve kelime bilgisi gibi faydalı bilgilere erişim sağlayan bir açılır menüden oluşur. Bununla birlikte, bildirilen bu vakalar, analiz edildiğinde, yakın gelecekte platformun olası bir olayı öngörmesini ve otomatik bir şekilde tepki vermesini sağlayacak tanımlanabilir davranış kalıplarına yol açan bir veri tabanı oluşturacaktır.

Google, yapay zekâdaki en son büyük başarının arkasındaki şirkettir. “Alpha Go”, ilk genel istihbarat programı olarak kabul edilir. Google tarafından 2014 yılında satın alınan yapay zekâ şirketi “Deep Mind” tarafından geliştirilen program, yalnızca bir hareket kaydını analiz ederek öğrenmesine izin veren makine öğrenimini kullanmakla kalmaz, aynı zamanda kendine karşı oynayarak öğrenilen stratejileri tasarlamasına izin veren güçlendirilmiş öğrenmeyi de entegre eder. Geçen yıl bu program, insan zekâsı tarafından şimdiye kadar yaratılmış en karmaşık oyun olarak kabul edilen Go'nun en büyük ustası Lee Sedol'u yendi. Bu gerçek, yalnızca yapay zekâyı çevreleyen tanıtım yutturmacasına katkıda bulunmakla kalmadı, aynı zamanda şirketi bu yeni teknolojik çerçevenin başına getirdi. Web arama motorlarının evrimine damgasını vuran değişikliklere öncülük eden Google, şimdi bu ortamla ilişkimizi yöneten paradigmayı değiştirecek bir yapay zekâ uygulaması öneriyor. Bu değişiklik, metni Larry Page ve Sergey Brin tarafından Google asistanını tanıtan Google CEO'su Sundar Pichai'ye atanan bu yıl yatırımcılara yönelik mektupta tanıtıldı.

Google, girdiğimiz arama terimlerini otomatik olarak tamamlamak ve düzeltmek için arama motoruna makine öğrenimi uygular. Bu amaçla, kendi çevirmeni ve ses tanıma özelliğini geliştirmesine ve Allo'yu, bir konuşma arayüzü oluşturmasına olanak tanıyan bir teknoloji olan doğal dil işlemeyi kullanır. Bilgisayarlı görüntü, görsel arama hizmetini ortaya çıkardı ve yeni Google Fotoğraflar uygulamasının, görsellerimizi önceden etiketlemeye gerek kalmadan sınıflandırmasını sağlayan şey de budur. Diğer yapay zekâ uygulamaları, Perspective'in çevrimiçi tacizi ve kötüye kullanımını azaltmak ve hatta veri sunucusu çiftliklerinin enerji maliyetini azaltmak için toksik yorumları analiz etmesine ve raporlamasına olanak tanır.

Google asistanı, arama sonuçları sayfasını konuşma arayüzü yerine değiştirerek platform hakkında bilgi edinmenin yeni bir yolunu temsil edecek. Bunda, akıllı bir temsilci

bağlamımızı, durumumuzu ve ihtiyaçlarımızı anlamak için tüm hizmetlere erişecek ve yalnızca bir seçenekler listesi değil, sorularımıza yanıt olarak bir eylem de üretecektir. Bu şekilde, Google artık bir gösteri, yayın zamanı, yeri ve biletlerin satışı ile ilgili bilgilere erişim sağlamak yerine, giriş biletlerini satın alacak ve gösteriyi takvimimize programlayacak entegre bir hizmet sağlayacaktır. Bu asistan, günlüğümüzü organize edebilecek, ödemelerimizi ve bütçelerimizi yönetebilecek ve cep telefonlarımızı tüm hayatımızın uzaktan kumandalarına dönüştürmeye katkıda bulunacak diğer birçok şeyi yapabilecek.

Makine öğrenimi, kullanımla birlikte gelişen otonom sistemler üreten verilerin analizine dayanır. Bu sistemler, tüm internet ortamını fetheden hızlı bir ilerlemeyle kendi inovasyon ekosistemlerini oluşturuyor. Spotify'nın öneri sistemini yöneten akıllı algoritmalar, Shazam uygulamasının şarkıları dinlemesine ve tanınmasına izin veren şeydir ve bunları yalnızca ürünlerini önermek ve dağıtmak için değil, aynı zamanda kullanıcılarının zevkine uygun prodüksiyonu planlamak, diziler ve filmler sunmak için kullanan Netflix'in başarısının arkasındadır.

Veri üreten cihazların sayısı arttıkça, her yere yapay zekâ sızıyor. Amazon bunu yalnızca öneri algoritmalarında değil, aynı zamanda lojistiğinin yönetiminde, ürünlerini taşıyabilen ve teslim edebilen otonom araçların oluşturulmasında da kullanıyor. Ulaşım paylaşım uygulaması Uber, bunları sürücülerin ve kullanıcıların itibarının profilini çıkarmak, onları eşleştirmek, rota önermek ve değişken sistemi içinde fiyatları hesaplamak için kullanır. Bu etkileşimler, şirketin otonom aracının üretiminde kullandığı bir veri tabanı üretir.

2001 yılında GPS sisteminin araçlara uygulanmasından bu yana, yeni sensörlerin geliştirilmesiyle birlikte büyük bir navigasyon veri tabanı üretildi ve bu da Google'ın şu anda 500.000 km'yi kazasız bir şekilde kat eden otonom bir araç oluşturmasını mümkün kıldı ve ticarileşmesini Waymo adı altında duyurdu.

Sağlık için tasarlanmış algoritmalar ve robotlar, Google Home ve Amazon Echo gibi evlerimiz için asistanlarda ve hayati belirtilerimiz hakkında veri toplayan giyilebilir cihazlarda uygulanmaktadır ve teşhis görüntülerinin ve tıbbi vaka geçmişlerinin dijitalleştirilmesiyle birlikte tahmine dayalı bir uygulamaya yol açmaktadır. Ayrıca, güvenlik kameralarının ve polis kayıtlarının çoğaltılması, akıllı algoritmaların suç tahminine ve adli kararların alınmasına olanak sağlıyor.

Sosyal etkileşimlerimizin gerçekleştiği algoritmik ortam, akıllı ve özerk hale geldi, aynı zamanda sosyal ağlardan tüm çevremize yayılmak için göç ederken, davranışlarımızı tahmin etme ve kontrol etme kapasitesini artırdı. Yapay zekâdaki yeni patlama, bu teknolojik dokuyu, algoritmaların çevreden öğrenmesini sağlayan veriler tarafından sürdürülen entellektüel süreçlerin mantıksal tanımından yola çıkaran bir paradigma değişikliğinden kaynaklanmaktadır.

Yapay zekâ, daha fazla kişiselleştirme ve makinelerle daha kolay ve daha entegre bir ilişki vaat ediyor. Ulaşım, sağlık, eğitim veya güvenlik gibi alanlarda refahımızı korumak, olası risklere karşı bizi uyarmak ve istendiğinde hizmet almak için kullanılır. Bununla birlikte, bu algoritmaların uygulanması, bu sistemin kırılganlığı konusunda uyaran bazı skandal olaylara yol açmıştır. Bunlar arasında dramatik Tesla yarı otomatik araç kazası, Facebook ve Twitter gibi ağlarda yanlış haberlerin yayılması, Microsoft tarafından geliştirilen ve kullanıcılarla etkileşim halinde öğrenmek için Twitter platformunda yayınlanan Tay botu ile yapılan başarısız deneyler yer alıyor. Saldırgan yorumları nedeniyle 24 saatten daha kısa bir sürede geri çekilmesi gerekiyordu. Google Fotoğraflar'da Afro-Amerikalıların "goriller" olarak etiketlenmesi, Google'ın kadınlara erkeklere kıyasla üst düzey iş ilanları gösterme olasılığının daha düşük olduğunun teyidi veya Afro-Amerikalı suçluların Kafkasyalılarından daha sık suçlu potansiyeli olarak sınıflandırılması. Diğer sorunların yanı sıra, bu algoritmaların ayrımcı gücünü, ortaya çıkan davranış kapasitelerini ve insanlarla işbirliğindeki zorluklarını göstermiştir.

Bu ve diğer sorunlar, öncelikle Makine Öğreniminin doğasından, büyük verilere bağımlılığından, büyük karmaşıklığından ve öngörme kapasitesinden kaynaklanmaktadır. İkincisi, bu uygulamalı yöntemler Apple, Facebook, Google, IBM ve Microsoft gibi birkaç şirkette yoğunlaşmaktadır. Sorunların bulunduğu sosyal uygulamalar ve eşitlikçi erişimi garanti etmenin zorluğu da önemli faktördür. Bu algoritmalar yavaş yavaş toplumun kritik yapısına nüfuz ettikçe gerçekleşecek değişikliklere karşı dayanıklılık için stratejiler oluşturulmalıdır.

Algoritmaların tarafsız olmaması, büyük verilere bağımlılıklarından kaynaklanmaktadır, veritabanları tarafsız değildir ve toplandıkları donanımın doğasında bulunan önyargıları, derlenme amaçlarını ve eşit olmayan veri ortamını sunar. Ne tüm kentsel alanlarda, ne de tüm sosyal sınıflar ve olaylar açısından aynı veri yoğunluğu yoktur. Bu verilerle eğitilmiş algoritmaların uygulanması, kültürümüzde var olan önyargıları bir virüs gibi yayarak kısır döngülere ve toplum kesimlerinin marjinalleşmesine neden olabilir. Bu sorunun tedavisi, kapsayıcı veri tabanlarının üretilmesi ve bu algoritmaların sosyal değişime yöneliminde bir odak değişikliğini içerir.

Kitle kaynak kullanımı, daha adil veritabanlarının oluşturulmasını destekleyebilir, her durumda hangi verilerin hassas olduğunu değerlendirmek için işbirliği yapılabilir, olumsuzluklar ortadan kaldırılmaya devam edebilir ve uygulamaların tarafsızlığı sürekli test edilebilir. Bu anlamda Columbia, Cornell ve Saarland üniversitelerinden bir ekip, bir programda oluşabilecek haksız ilişkilendirmeleri arayan FairTest aracını oluşturdu. Ayrıca, sosyal değişime yönelik algoritmalar, kültürümüzde mevcut olan önyargıların tespit edilmesine ve ortadan kaldırılmasına katkıda bulunur. Boston Üniversitesi, Microsoft Research ile işbirliği içinde, İngilizce dilindeki önyargıların, özellikle de metnin otomatik

olarak sınıflandırılması için çeviri ve arama motorları gibi birçok uygulamada kullanılan Word2vec veritabanında ortaya çıkan haksız çağrışımların tespiti için algoritmaların kullanıldığı bir proje yürütülmüştür. Bu veri tabanından önyargıyı ortadan kaldırmak, onu kültürümüzden çıkarmaz, ancak yinelenen bir şekilde çalışan uygulamalar aracılığıyla yayılmasını önler.

Diğer sorunlar, yalnızca bu algoritmaların onları uygulayan şirketlerin mülkiyeti olarak kabul edilmesi ve korunmasından değil, aynı zamanda karmaşıklığından da kaynaklanan şeffaflık eksikliğinden kaynaklanmaktadır. Bununla birlikte, bu açıklayıcı algoritmaları oluşturan süreçlerin geliştirilmesi, bunların tıbbi, yasal veya askeri karar verme süreçlerine uygulanması halinde hayati bir öneme sahiptir; Bu anlamda Amerikan Savunma İleri Araştırma Projeleri Ajansı (DARPA: Defense Advanced Research Projects Agency), “Açıklanabilir Yapay Zekâ Programını” başlattı. Bu program, sınıflandırmalarıyla ilgili olduğu düşünülen bir görüntünün alanlarını vurgulayarak veya sonucu örnekleyen bir veritabanı örneği göstererek, akıl yürütmelerinin bir açıklamasını içerebilen yeni derin bilgi sistemlerini araştırır. Ayrıca, doğal dilde görselleştirmeler ve açıklamalar yoluyla, verilerle derin öğrenme sürecini daha açık hale getiren arayüzler geliştirirler. Bu prosedürlerin bir örneği, Google deneylerinden biri olan ve 2015 yılında üstlenilen Derin Rüya, derin öğrenmeye dayalı bir görüntü tanıma sisteminin değiştirilmesinden oluşuyordu, böylece fotoğraflarda bulunan nesneleri tanımlamak yerine onları değiştirdi. Bu ters işlem, programı işlevsel çerçevesinin dışında çalışmaya ve iç işleyişini ortaya çıkarmaya zorlayan bir yapı söküm süreci aracılığıyla, programın görüntüleri tanımlamak için seçtiği özelliklerin görselleştirilmesi için düşsel görüntülerin yaratılmasına izin verir.

Son olarak, bu sistemlerin tahmin kapasitesi, kontrol kapasitelerinde bir artışa yol açar. Ağ bağlantılı teknolojilerin kullanımından kaynaklanan gizlilik sorunları iyi bilinmektedir, ancak yapay zekâ önceki kararlarımızı analiz edebilir ve gelecekteki olası faaliyetlerimizi tahmin edebilir. Bu, sisteme, uygulamanın sosyal kontrolünü ve sorumlu kullanımı gerektiren kullanıcıların davranışlarını etkileme kapasitesi verir.

Ex Machina, bize yapay zekâyı çevreleyen, yeteneklerimizi aşan ve kontrolümüzden kaçan korkunun bir metaforunu sunuyor. Yapay zekânın insan evrimimizin gidişatını değiştirecek bir tekillik veya olay üretme olasılığı uzak olmaya devam ediyor. Ancak makine öğrenimindeki akıllı algoritmalar çevremizde yaygınlaşıyor ve önemli sosyal değişiklikler üretiyor, bu nedenle gereklidir. Tüm sosyal ajanların bu algoritmaların oluşturduğu süreçleri anlamalarına, tanımalarına ve uygulamalara katılmalarına izin veren stratejiler geliştirilmektedir.

Çocuk iken anne ve babamız bizden daha zekiydiler. Korkulacak bir durum yoktu. Çünkü onların amaçları ile çocuklarının amaçları aynıydı. Yapay zekâ, daha iyi bir toplum oluşturabilmek için yapılacaktır. Daha iyi, ilham verici bir gelecek yaratma fırsatı sunacaktır.

Yapay zekâ bir araçtır; bir doktora, bir mühendise, işinde daha iyi olabilmek için yardımcı olacak bir araçtır. Bilim insanının keşif sürecini hızlandıracak bir araçtır. Yapay zekâ, her şeyi bizim için yapan çerçeveler içine sokulmamalıdır. Bize hükmeden, tüm kontrolü elimizden alan bir güç olarak da görülmemelidir. Teknoloji hedeflerimize ulaşmamızı sağlayan bir araçtır. Teknolojinin kime hizmet edeceğine biz karar vermeliyiz.

Yapay Zeka

Bilgisayar bilimi, bir disiplin olarak, algoritmaların teorik çalışmalarından hesaplama ve hesaplama sınırları çalışmalarına, donanım ve yazılım alanlarından bilgisayar sistemlerinin uygulanmasına ilişkin pratik ve teorik olmak üzere geniş bir alanı kapsar.

Bilgisayar biliminde önemli olduğunu düşünülen alanlar şöyle tanımlanmaktadır:

- Hesaplama teorisi
- Algoritmalar ve veri yapıları
- Programlama metodolojisi ve dilleri
- Bilgisayar elemanları ve mimarisi
- Gömülü Sistemler: Bilgisayarı oluşturan temel bileşenlerin tümü tek bir chip içindedir. Uygulamaya yönelik yazılım; yapay zeka algoritmaları; program yazan algoritmalar.

Veri Bilimi, sonuçlar çıkarmak amacıyla hesaplamalı matematik, istatistik, olasılık ve makine öğrenimi tekniklerini kullanarak ham verileri analiz etme bilimidir. Basit bir ifadeyle, veri bilimindeki bir ardışık düzen, “çeşitli kaynaklardan (anketler, geri bildirim, satın alma listeleri, oylar vb.) elde edilen verileri saklar ve analiz için kullanır.”



Data science pipeline in a simplified way

Ham veriler, bir ardışık düzen içinde farklı aşamalardan geçer:

- Verilerin Getirilmesi/Elde Edilmesi
- Verileri Temizleme/Temizleme
- Veri görüntülenmesi
- Verileri Modelleme
- Verileri Yorumlama
- Revizyon



Yapay zeka, insan davranışını taklit edebilen akıllı makine yaratmayı vurgulayan bilgisayar bilimi teknolojisidir. **Burada Akıllı makineler, insan gibi davranabilen, insan gibi düşünebilen ve aynı zamanda karar verebilen makineler olarak tanımlanabilir.** "Yapay" ve "insan yapımı düşünme yeteneği" anlamına gelen "Zeka" olmak üzere iki kelimedenden oluşur. İnsan yapımı bilgisayar sistemleri içeren bir otonom makineden bahsedilmektedir. Yapay zeka ile, bir görevi gerçekleştirmek için makineyi önceden programlamamız gerekmez; bunun yerine akıllı algoritmalar ile bir makine oluşturabiliriz ve o kendi kendine programla yapacak ve çalışacak.

Makineden ne istiyoruz. Bir işi mükemmel bir şekilde yapmasını istiyoruz. Öğrenerek performansının yükseltilmesi istenir. Bu işi yaparken, daha kaliteli yapacak bir metot öğrenirse ise, bize sorma yap....

Komplo senaryosu: Bizden daha zeki makineler yaratabiliyor olursak, o zaman bu makine nesli, yeniden daha akıllı makineler yaratabilecek mi? Bu, insan zekasının hızlı ve geri döndürülemez bir şekilde makine zekası tarafından çok geride bırakıldığı üstel bir duruma yol açacaktır. Sonuç olarak, yetki ve kontrolü kaybederiz. En iyi durumda, makinelerin kölesi oluyoruz; en kötü durumda, ihtiyaç fazlası olarak yok ediliriz. Makinelerin insanların yaşam alanlarını hapishaneye dönüştürebilir mi? Bir sınıfın elinde güç olarak kullanılabilir mi?

Neden Yapay Zeka?

Yapay zekanın amacı, insan davranışını taklit edebilen akıllı makineler yaratmaktır. Günümüz dünyasında karmaşık sorunları çözmek, rutin işleri otomatikleştirerek hayatımızı daha sorunsuz hale getirmek, insan gücünden tasarruf etmek ve daha birçok görevi yerine getirmek için yapay zekaya ihtiyacımız var.

İnsan Zekası, insanların sorunları düşünme, analiz etme ve çözme ve gelecekteki durumlara yardımcı olan geçmiş deneyimlerden öğrenme konusundaki entelektüel yeteneğidir. **Yapay zeka (AI), görevleri insanlarla aynı şekilde gerçekleştirmek için eğitilmiş bir makinede insan zekasının simülasyonudur.**

GPT-4 (Generative Pre-trained Transformer 4), OpenAI tarafından oluşturulan çok modlu büyük bir dil modelidir. GPT serisinin dördüncüsüdür. 14 Mart 2023'te yayınlanmıştır ve API aracılığıyla ve ChatGPT Plus kullanıcıları için kullanılabilir olacaktır. Microsoft, GPT kullanan Bing sürümlerinin resmi yayınlanmadan önce aslında GPT-4 kullandığını doğruladı.

Yetenekleri

- GPT-4 daha karmaşık ve nüanslı girdileri anlayabilir. OpenAI'ye göre, GPT-4 " **çeşitli profesyonel ve akademik kriterlerde insan seviyesinde performans sergiliyor.**
- GPT-4 çok modlu yeteneklere sahiptir. Sadece metin değil, aynı zamanda görüntü ve video gibi diğer ortamları da işleyebilir.
- GPT-4 yüksek derecede yönlendirilebilirlik sunar. OpenAI ayrıca, kullanıcıların modelin çıktısını istedikleri şekilde etkileyebilecekleri ve yönlendirebilecekleri konusunda iddialıdır.
- GPT-4 daha güvenli ve yararlı yanıtlar üretir. OpenAI, GPT-4'ün zararlı veya yanlış bilgileri azalttığını ve kullanıcıların ihtiyaçlarına uygun yanıtlar verdiğini söylüyor.
- GPT-4 performans iyileştirmeleri sunar. OpenAI, GPT-4'ün önceki sürümlerine göre daha hızlı ve verimli olduğunu belirtiyor.

OpenAI tarafından geliştirilen ve çığır açan bir dil modeli olarak genel kabul gören ChatGPT, ilk duyurulduğundan bu yana dikkate değer bir evrim geçirdi. GPT-4 mimarisine dayanan ChatGPT, mevcut en gelişmiş yapay zeka dil modellerinden biri haline geldi ve kullanıcıların sistemle doğal ve dinamik konuşmalar yapmasına olanak sağladı.

ChatGPT'nin yolculuğu, 2018'de orijinal GPT ile başlayan GPT serisinin piyasaya sürülmesiyle başladı. Bu model, Vaswani ve diğerleri tarafından tanıtılan bir sinir ağı modeli olan Transformer mimarisine dayanıyordu.

OpenAI'nin 2019'da piyasaya sürülen GPT-2'si, tutarlı ve bağlamsal olarak alakalı metin oluşturma becerisiyle büyük ilgi gördü, ancak bu modellerin gerçek potansiyeli, 2020'de GPT-3'ün piyasaya sürülmesine kadar ortaya çıkmadı.

GPT-3'ün dikkate değer dil anlama ve oluşturma yetenekleri, ChatGPT'nin geliştirilmesi için de oldukça büyük bir zemin hazırladı.

Bu 10 maddede serimizde gündemden düşmeyen GPT-4'ün doğuşu, başarısı ve geliştiriciler için ne ifade ettiğini ele alacağız.

GPT-3'ün doğrudan soyundan gelen ChatGPT, kullanıcılarla daha doğal, dinamik konuşmalar yapmak için GPT-3 modülünü farklı bir platform üzerinde entegre etti. Model, konuşma bağlamında gelişmiş yanıtlar üretmesine izin veren İnsan Geri Bildiriminden (RLHF) Güçlendirmeli Öğrenim kullanılarak günden güne geliştirildi. ChatGPT'nin ilk sürümleri, öncelikle araştırma ve deneyler için kullanılıyordu. Bununla birlikte OpenAI, geliştiricilerin ve araştırmacıların uygulama ve hizmetler oluşturmaları için API'ye sınırlı erişim sağlıyordu. Bu yüzden kullanıcıların ChatGPT'yi denemeyebilmesi için bir Waitlist beklemeleri gerekiyordu.

Teknoloji ve yapay zeka (AI) geliştikçe, bazı mesleklerin yapısı ve talebi değişir. GPT-4 ve benzeri AI teknolojileri, özellikle tekrarlayan ve düşük seviyeli analitik beceri gerektiren işleri otomatikleştirebilir. Bu, bazı işlerin ve mesleklerin risk altında olduğu anlamına gelir.

Ancak, AI aynı zamanda yeni iş olanakları ve sektörlerin gelişmesine de katkıda bulunabilir. Örneğin, AI ve teknoloji alanındaki işlerde, veri bilimi ve makine öğrenimi mühendisliği gibi rollerin talebi artmaktadır.

GPT-4 veya benzer AI sistemleri mesleğinizi elimizden alıp almayacağı, işinizin doğası ve ne kadar rutin ve öngörülebilir olduğuna bağlıdır. AI teknolojilerinin gelişmesiyle, insanlar daha yaratıcı, eleştirel düşünme ve insanlar arası iletişim becerilerine dayalı işlerde daha çok yer bulabilirler. Bu nedenle, AI sistemlerine uyum sağlamak ve teknolojiyi avantaja dönüştürmek için becerilerinizi geliştirmeye odaklanmanız önemlidir.

Sonuç olarak, GPT-4 ve benzeri AI sistemleri bazı mesleklerde işleri otomatikleştirebilir, ancak aynı zamanda yeni iş fırsatları da yaratabilir. AI'ye uyum sağlamak ve sürekli öğrenmeye odaklanmak, gelecekteki iş güvenliğinizi ve başarınızı sağlamak için önemlidir.

Yapay zeka türleri

Yapay zeka, yeteneklere ve işlevlere göre farklı türlere ayrılabilir.

Yeteneklere Göre:

- Zayıf AI veya Dar AI: Zayıf AI, zeka ile bazı özel görevleri yerine getirme yeteneğine sahiptir. Siri, Zayıf AI'nın bir örneğidir.
- Genel AI: Herhangi bir entelektüel görevi bir insan gibi verimlilikle gerçekleştirebilen akıllı makineler.
- Güçlü AI: İnsanlardan daha iyi olacak ve insan zekasını aşacak makineyi içeren varsayımsal bir kavramdır.

İşlevlere Göre:

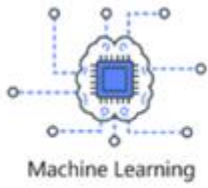
- Reaktif Makineler: Tamamen reaktif makineler, AI'nın temel türleridir. Bunlar mevcut işlere odaklanır ve önceki işleri saklayamaz.
- Sınırlı Bellek: Adından da anlaşılacağı gibi, geçmiş verileri veya deneyimleri sınırlı bir süre boyunca saklayabilir. Kendi kendini süren araba, bu tür AI türlerine bir örnektir.
- Zihin Teorisi: Gerçek dünyada insan duygularını, insanları vb. anlayabilen gelişmiş yapay zekadır.
- Öz Farkındalık: Öz Farkındalık Yapay zeka, insanlara benzer kendi bilincine, duygularına sahip olacak Yapay Zekanın geleceğidir.

AI'nın farklı alanları/Alt kümeleri

AI birçok alanı veya alt grubu kapsar ve bazı ana alanlar aşağıda verilmiştir: Mükemmel insan yok. Aslında varlar. Nerdeler? Mezarlıktalar... Mükemmel olma yolunda çaba sarfedenler var.

- Makine Öğrenimi
- Derin Öğrenme
- Sinir Ağı
- Uzman Sistem
- Bulanık Mantık
- Doğal Dil İşleme
- Robotik
- Konuşma Tanıma.

Yapay Zekanın farklı alanları



Machine Learning



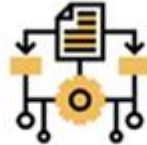
Neural Networks



Robotics



Expert Systems



Fuzzy Logic



Natural Language
Processing

- **Makine Öğrenimi:** Açıkça programlanmadan kendi kendilerine birkaç numara öğrenebilmeleri için bilgisayarları veri besleyerek harekete geçirme bilimidir.
- **Sinir Ağları:** İnsan beynine göre modellenmiş bir dizi algoritma ve tekniktir. Sinir Ağları, karmaşık ve gelişmiş makine öğrenimi sorunlarını çözmek için tasarlanmıştır.
- **Robotik:** Robotik, robotların farklı dallarını ve uygulamalarını içeren yapay zekanın bir alt kümesidir. Bu Robotlar, gerçek dünya ortamında hareket eden yapay etkenlerdir. Bir AI Robot, çevresindeki nesneleri algılayarak, öğrenerek, öğretilerek, hareket ettirerek ve ilgili eylemleri gerçekleştirerek manipüle ederek çalışır.
- **Uzman Sistemler:** Bir uzman sistem, bir insanın karar verme yeteneğini taklit eden bir bilgisayar sistemidir. Belirli bir alanda uzman bilgi ve deneyime sahip bir insan veya kuruluşun yargı ve davranışlarını simüle etmek için yapay zeka (AI) teknolojilerini kullanan bir bilgisayar programıdır.
- **Bulanık Mantık Sistemleri:** Bulanık mantık, modern bilgisayarın dayandığı olağan "doğru veya yanlış" (1 veya 0) boole mantığından ziyade "doğruluk derecelerine" dayalı bir hesaplama yaklaşımıdır. Bulanık mantık Sistemleri, kesin olmayan, bozuk, gürültülü giriş bilgilerini alabilir.
- **Doğal Dil İşleme:** Doğal Dil İşleme (NLP), sorunları çözmek için yararlı içgörüler elde etmek için doğal insan dilini analiz eden Yapay Zeka yöntemini ifade eder.

AI için kullanılan programlama dilleri

Aşağıda, Yapay Zekanın geliştirilmesi için yaygın olarak kullanılan ilk beş programlama dili verilmiştir:

- Python
- Java Script
- C++
- Lisp
- Prolog

Yukarıdaki beş dil arasında Python, sadeliği ve Numpy, Pandas vb. gibi birçok kütüphanesinin kullanılabilirliği nedeniyle AI geliştirme için en çok kullanılan dildir.

Yapay zekada genellikle hangi programlama dili kullanılmaz ve neden?

Perl Programlama dili, betik dili olduğu için AI için yaygın olarak kullanılan bir dil değildir.

AI geliştirme için yazılım platformları

- Google Cloud AI platformu
- Microsoft Azure AI platformu
- IBM Watson
- TensorFlow
- Infosys Nia
- Rainbird
- Dialogflow

Yapay zekadaki akıl faktörü

Akıllı faktörü, çevresini ve değişimleri sensörler aracılığıyla algılayan ve amacına ulaşmak için aktüatörleri kullanarak üzerinde hareket eden herhangi bir otonom makine olabilir.

AI'daki bu Akıllı araçlar aşağıdaki uygulamalarda kullanılır:

- Arama Motoru gibi Bilgi Erişimi ve Gezinmeler
- Tekrarlayan Faaliyetler
- Alan Uzmanları
- Sohbet robotları vb.

Yapay zekadaki Turing testi

Turing testi, Yapay zeka alanındaki popüler zeka testlerinden biridir. Turing testi 1950 yılında Alan Turing tarafından tanıtıldı. **Bir makinenin insan gibi düşünüp düşünemeyeceğini belirlemek için yapılan bir testtir.** Bu teste göre, bir bilgisayarın yalnızca belirli koşullar altında insan tepkilerini taklit edebiliyorsa zeki olduğu söylenebilir.

Bu testte üç oyuncu yer alır, ilk oyuncu bilgisayar, ikinci oyuncu insan yanıtlayıcı ve üçüncü oyuncu insan sorgulayıcıdır ve sorgulayıcının sorulara dayanarak makineden hangi yanıtın geldiğini bulması gerekir ve cevaplar. **Makinenin zekasını test etmek için Turing testi değerlendirmesi kullanılır.**

Bir makinenin zekasını test etmek için kullanılan değerlendirme

Yapay zekada (AI), Turing Testi, bir bilgisayarın insan gibi düşünüp düşünemeyeceğini belirlemek için bir araştırma yöntemidir.

Yapay zekada bilgisayarla görme

Bilgisayarla görme, bilgisayarları görsel dünyadan görüntüler gibi yorumlayabilmeleri ve bilgi alabilmeleri için eğitmek için kullanılan bir Yapay Zeka alanıdır. Bu nedenle, bilgisayarla görme, görüntü işleme, nesne algılama vb. gibi karmaşık sorunları çözmek için AI teknolojisini kullanır.

Yapay zeka ile ilgili bazı yanlış anlamalar

Yapay zeka evrimine başladığından beri birçok yanlış anlama var. Bu kavram yanlışlarından bazıları aşağıda verilmiştir:

- AI insana ihtiyaç duymaz: AI hakkındaki ilk yanlış kanı, insana ihtiyaç duymamasıdır. Ancak gerçekte, AI tabanlı her sistem bir yerde insanlara bağımlıdır ve kalacaktır. Örneğin, veriler hakkında bilgi edinmek için insan tarafından toplanan verileri gerektirir.
- AI insanlar için tehlikelidir: AI insanlar için doğası gereği tehlikeli değildir ve yine de süper AI veya insanlardan daha zeki olan güçlü AI'ya ulaşmamıştır. Herhangi bir güçlü teknoloji, kötüye kullanılmadığı takdirde zararlı olamaz.
- AI zirve aşamasına ulaştı: Yine de, AI'nın zirve aşamasından çok uzagız. Zirveye ulaşmak için çok uzun bir yolculuk gerekecek.
- AI işinizi alacak: AI'nın işlerin çoğunu alacağı en büyük kafa karışıklıklarından biridir, ancak gerçekte bize yeni işler için daha fazla fırsat sunuyor.
- AI yeni bir teknolojidir: Bazı insanlar bunun yeni bir teknoloji olduğunu düşünse de, bu teknoloji aslında ilk olarak 1840 yılında bir İngiliz gazetesi aracılığıyla düşünülmüştür.

Yapay zekada bilgi temsili

Bilgi temsili, AI faktörlerinin düşünmesiyle ilgilenen AI'nın bir parçasıdır. AI faktörlerine gerçek dünya hakkındaki bilgileri temsil etmek için kullanılır, böylece AI'daki karmaşık sorunları çözmek için bu bilgileri anlayabilir ve kullanabilirler.

AI sisteminde aracıya temsil edilen aşağıdaki Bilgi öğeleri:

- Nesneler -IoT (Gömülü sistem, Algılayıcılar, Haberleşme 5G/6G, Yazılımlar, Akıllı Algoritmalar)
- Etkinlikler
- Performans
- Meta Bilgi
- Gerçekler
- Bilgi tabanı

Yapay zekada çeşitli bilgi temsili teknikleri

Bilgi temsil teknikleri aşağıda verilmiştir:

- Mantıksal Temsil
- Semantik Ağ Temsili
- Çerçeve Gösterimi
- Üretim Kuralları

Yapay zekanın büyük etkisi olduğu farklı alanlar

AI'nın büyük etkisi olduğu bazı alanlar şunlardır:

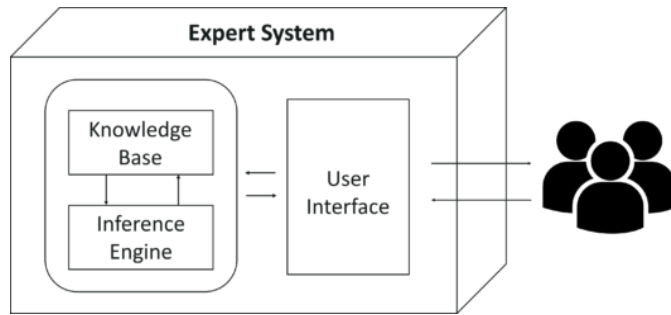
- Otonom Ulaşım
- Yapay zeka tarafından desteklenen eğitim sistemi.
- Sağlık
- Tahmine Dayalı Polislik
- Uzay Keşfi
- Eğlence vb.

Yapay Zekada Kavram Oluřturma Mantiđı

Yapay zekada kavram oluřturma mantiđı, bazı mantıksal kuralları uygulayarak veri tabanından yeni bilgiler tureten akıllı bir sistemin parçasıdır. Esas olarak iki modda çalışır:

- Geriye Zincirleme: Hedefle başlar ve hedefi destekleyen gerçekleri çıkarmak için geriye doğru ilerler.
- İleri Zincirleme: Bilinen gerçeklerle başlar ve yeni gerçekleri öne sürer.

Uzman Sistemlerin bileřenleri



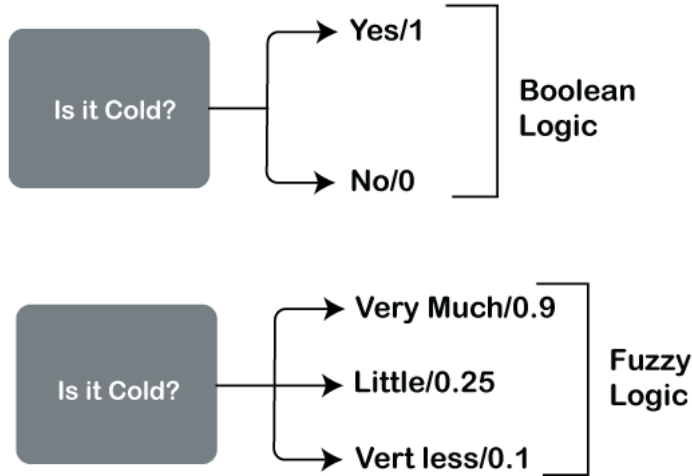
- Bilgi (**Knowledge**) tabanı: Özel ve yüksek kaliteli bilgi içerir.
- Çıkarım (Inference) Mantiđı: Belirli bir çözüme ulaşmak için bilgi tabanından bilgiyi alır ve manipüle eder.
- Kullanıcı arayüzü (User Interface): Kullanıcı arayüzü, kullanıcı ile Uzman Sistemin kendisi arasında etkileřim sağlar.

Bir uzman sistem temel olarak üç bileřen içerir:

- Kullanıcı Arayüzü (User Interface): Bir problemin çözümünü bulmak için bir kullanıcının uzman sistemle etkileřime girmesini veya iletiřim kurmasını sağlar.
- Çıkarım Mantiđı (Inference Engine): Uzman sistemin ana iřlem birimi veya beyni olarak adlandırılır. Bundan bir sonuç çıkarmak için bilgi tabanına farklı çıkarım kuralları uygular. Sistem, bir çıkarım motoru yardımıyla KB'den bilgileri çıkarır.
- Bilgi Bankası (Knowledge Base): Bilgi tabanı, alana özgü ve yüksek kaliteli bilgiyi depolayan bir tür depolama alanıdır.

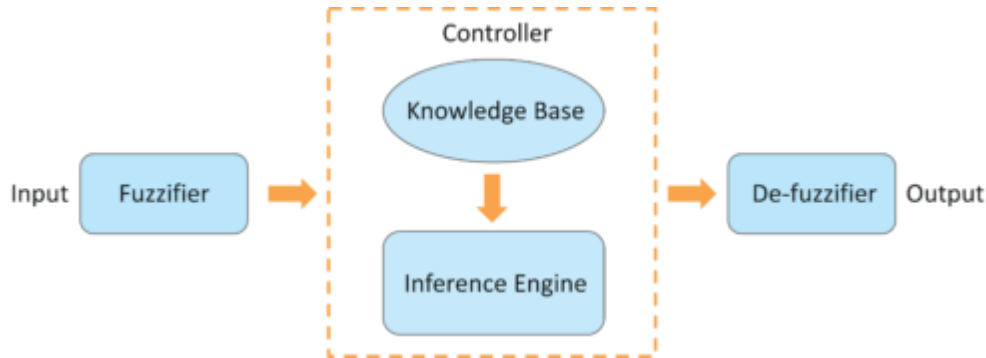
Bulanık mantık

Bulanık mantık, yapay zekaya uygulanan ve insan muhakemesine benzeyen bir akıl yürütme yöntemidir. Burada "bulanık" kelimesi net olmayan şeyleri tanımlar, durumun Doğru mu Yanlış mı olduğuna karar vermenin zor olduğu durumları ifade eder. Evet ile HAYIR arasında meydana gelen tüm olasılıkları içerir. Aşağıdaki şema, bulanık mantık ile Boolean mantığı arasındaki farkı göstermektedir.



İnsan muhakemesini andırdığı için sinir ağlarında kullanılabilir.

Bulanık Mantık mimarisi



- Bulanıklaştırma (**Fuzzification**) Modülü – Sistem girdileri, girdileri bulanık kümelere dönüştüren Bulanıklaştırıcıya beslenir.
- Bilgi (**Knowledge**) Tabanı – Uzmanlar tarafından sağlanan EVET-HAYIR kuralları gibi analitik ölçüleri depolar.
- Çıkarım Mantığı (**Inference**) – Girdiler ve IF-THEN kuralları üzerinde bulanık çıkarımlar yaparak insan akıl yürütme sürecini simüle eder.
- Durulaştırma (**Defuzzification**) Modülü – Çıkarım motoru tarafından elde edilen bulanık kümeyi kesin bir değere dönüştürür.

Oyun teorisinde Yapay Zeka

Oyun teorisi, iki veya daha fazla rasyonel oyuncu arasındaki olası etkileşimlerin bir modelini oluşturan mantıksal ve bilimsel çalışmadır. Burada rasyonel, her oyuncunun diğerlerinin de aynı derecede rasyonel olduğunu ve aynı düzeyde bilgi ve anlayışa sahip olduğunu düşünmesi anlamına gelir. Önyargılara dikkat edilmelidir. Oyun teorisinde, oyuncular çok etmenli bir durumda verilen seçenekler kümesiyle ilgilenir, bu, bir oyuncunun seçiminin diğer veya rakip oyuncuların seçimini etkilediği anlamına gelir.

Oyun teorisi ve yapay zeka birbiriyle çok ilgili ve kullanışlıdır. **Yapay zekada oyun teorisi, birden fazla oyuncunun bir hedefe ulaşmak için birbiriyle etkileşime girmeye çalıştığı çok etmenli ortamda gereken bazı temel yetenekleri etkinleştirmek için yaygın olarak kullanılır.**

Poker, Satranç vb. gibi farklı popüler oyunlar, **kuralları belirlenmiş mantıksal oyunlardır**. Bu oyunları çevrimiçi veya dijital olarak, örneğin Mobil, dizüstü bilgisayar vb.'de oynamak için, bu tür oyunlar için algoritmalar oluşturmak gerekir. Ve bu algoritmalar yapay zeka yardımıyla uygulanmaktadır.

Oyun teorisi ve AI ilişkisi

“Yapay zeka (AI) ve derin öğrenme sistemleri bağlamında, bir hedefe ulaşmak için farklı AI programlarının etkileşime girmesi veya rekabet etmesi gereken çok etmenli ortamlarda gerekli olan bazı temel yetenekleri etkinleştirmek için oyun teorisi esastır.”



Minimax Algoritması nedir? Bir Minimax probleminde yer alan terminolojileri açıklayın.

Minimax, diğer oyuncunun da en iyi şekilde oynadığını varsayarak bir oyuncu için en uygun hamleyi seçmek için kullanılan özyinelemeli bir algoritmadır.

Bir oyun, aşağıdaki bileşenlerle bir arama problemi olarak tanımlanabilir:

- **Oyun Ağacı:** Tüm olası hareketleri içeren bir ağaç yapısı.
- **Başlangıç durumu:** Başlangıç konumu ve kimin hamlesi olduğunu gösterir.
- **Ardıl işlevi:** Bir oyuncunun yapabileceği olası doğru hamleleri tanımlar.
- **Terminal durumu:** Oyun bittiğinde tahtanın konumudur.

- **Fayda işlevi:** Bir oyunun sonucu için sayısal bir değer atayan bir işlevdir.

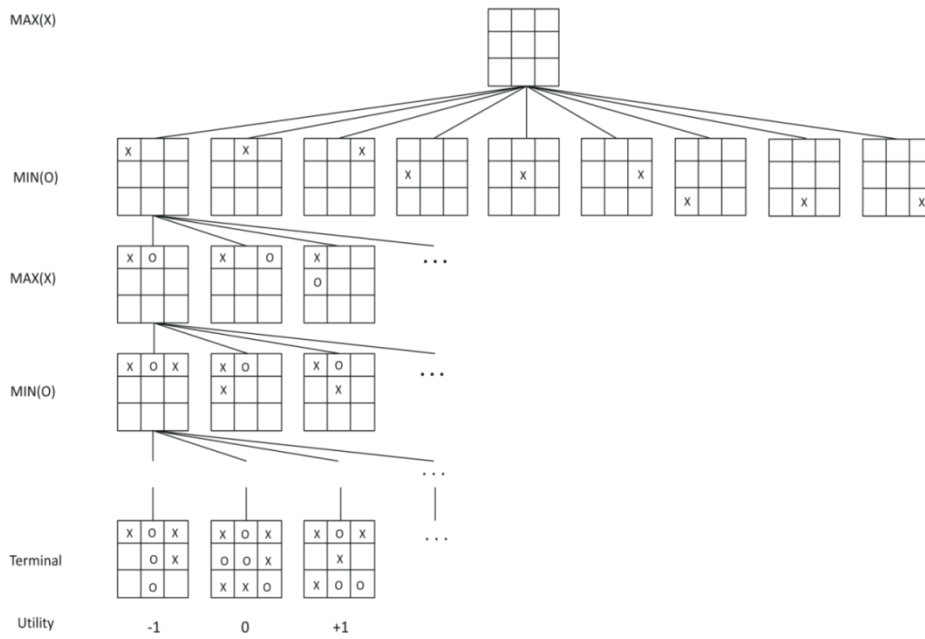
Tic-Tac-Toe Game kullanarak Minimax algoritmasının çalışmasının gösterilmesi

Bir oyunda yer alan iki oyuncu vardır:

- **MAKS:** Bu oyuncu mümkün olan en yüksek puanı almaya çalışır.
- **MIN:** MIN mümkün olan en düşük puanı almaya çalışır.

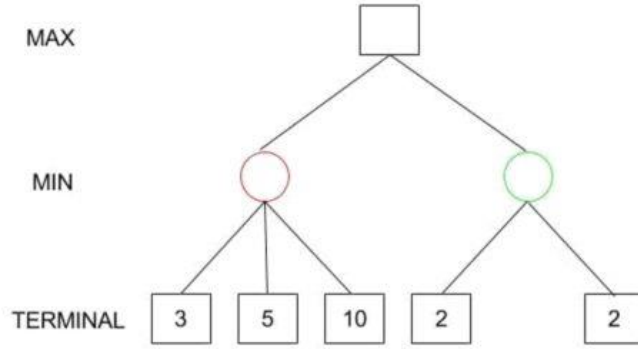
Minimax algoritmasını kullanan bir Tic-Tac-Toe oyunu için aşağıdaki yaklaşım benimsenmiştir.

Adım 1: İlk olarak, oyunun mevcut konumundan başlayarak terminal durumlarına kadar tüm oyun ağacı oluşturulur.



Adım 2: Tüm terminal durumları için yardımcı program değerlerini almak için yardımcı program işlevini uygulanır.

Adım 3: Terminal düğümlerinin yardımcı programlarının yardımıyla daha yüksek düğümlerin yardımcı programları belirlenir. Örneğin, aşağıdaki şemada, karelerde yazılmış terminal durumları için yardımcı programlar var.



Terminalin üzerindeki katmanın sol düğümü(kırmızı) için yardımcı programı hesaplayalım:

$\text{MIN}\{3, 5, 10\}$, yani 3.

Bu nedenle, kırmızı düğüm için yardımcı program 3'tür.

Benzer şekilde, aynı katmandaki yeşil düğüm için:

$\text{MIN}\{2,2\}$, yani 2.

Adım 4: Fayda değerlerini hesaplayın.

Adım 5: Sonunda, yedeklenen tüm değerler ağacın köküne ulaşır. Bu noktada MAX, en yüksek değeri seçmelidir:

yani, 3 olan $\text{MAX}\{3,2\}$.

Bu nedenle, MAX için en iyi açılış hareketi sol düğümdür (veya kırmızı olan).

Özetlemek,

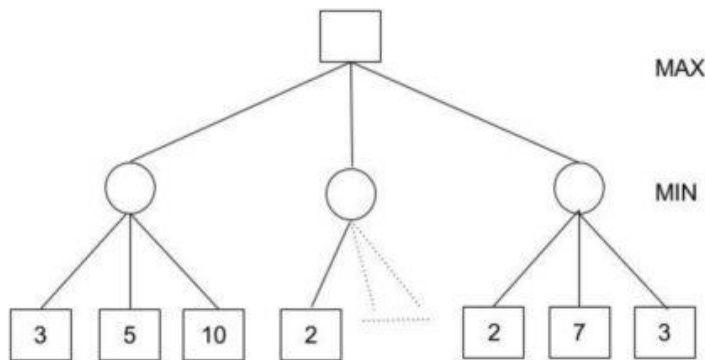
Minimaks Kararı = $\text{MAKS}\{\text{MIN}\{3,5,10\}, \text{MIN}\{2,2\}\}$

= $\text{MAKS}\{3,2\}= 3$

Minimax tabanlı bir oyunu optimize etmek

Alfa-beta Budama

Standart bir minimax algoritmasına alfa-beta budama uygularsak, standart olanla aynı hareketi döndürür, ancak muhtemelen nihai kararı etkilemeyen tüm düğümleri kaldırır.



Yapay Zeka Modelleme: Hesaplamaya Dayalı Yorumlamada İlerleyen Zeka

Yapay zeka modelleri (makine öğrenimi ve derin öğrenme) mantıksal çıkarımı ve karar almayı otonomlaştırmaya yardımcı olur.

Yapay Zeka Modelleme Nedir?

Veri toplama ve veri hazırlamanın ardından, veri bilimindeki birsonraki aşama, gelişmiş hesaplama yöntemlerini desteklemek için akıllı makine öğrenimi modellerinin oluşturulmasını içerir. Bu modeller, verilerdeki kalıpları tanımlamak ve insan uzmanlığını taklit edecek şekilde sonuçlar çıkarmak için doğrusal veya lojistik regresyon gibi çeşitli algoritma türlerini kullanır.

Kısacası, yapay zeka modelleme, üç temel adımı izleyen bir karar verme sürecinin oluşturulmasıdır:

- **Modelleme:** İlk adım, verileri yorumlayan ve bu verilere göre kararlar veren algoritmaları kullanan bir yapay zeka modeli oluşturmaktır. **Başarılı bir yapay zeka modeli, herhangi bir kullanım durumunda insan gibi otonom davranış geliştirebilir.**
- **Yapay zeka model eğitimi:** İkinci adım, yapay zeka modelini eğitmektir. Çoğu zaman eğitim, yinelenmeli test döngülerinde yapay zeka modeli aracılığıyla büyük miktarda verinin işlenmesini ve modelin beklediği ve istendiği gibi davrandığından emin olmak ve doğruluğu sağlamak için sonuçların kontrol edilmesini içerir. **Modelleme parametreler ve hiperparametreler bulunmaktadır. Parametreler modelin öğrenerek belirlediği değerlerdir. Hiperparametreler ise insan uzmanlığının, tecrübesinin aktarıldığı kesin değelerdir; değer aralığı da olabilir. Mühendisler bu süreçte yapay zeka modeli öğrendikçe onu değiştirmek ve geliştirmek için yetenek kazanırlar. Bu aşamada simülasyon (Benzerini yazılımsal oluşturma) ve Optimizasyon (En iyileme – hassasiyet bulma) çalışmaları yapılır.**
- **Çıkarım:** Üçüncü adım, çıkarım olarak bilinir. Bu adım, yapay zeka modelinin genellikle mevcut verilere dayanarak mantıksal sonuçlar çıkardığı gerçek dünyadaki kullanım durumuna açılmasını ifade eder.

Makine Öğrenimi ve Derin Öğrenme

Yapay Zeka ya da Makine Öğrenimi, yüksek hesaplama, depolama, veri güvenliği veri haberleşmesi ve ağ gereksinimlerine sahip karmaşık bir süreçtir.

Yapay zeka, makinelerin veya bilgisayarların insan gibi mevcut verilere dayanarak karar verme kapasitesini taklit ettiği herhangi bir yöntemi ifade eden bir genel terimdir. **Makine öğrenimi, otomatikleştirilmiş görevleri mümkün kılmak için özellikle Yapay Zekanın algoritmaları formunda uygulanmasıdır. Makine öğreniminin önemli bir özelliği, daha fazla veriyi kullandıkça öğrenmesi ve zamanla daha iyi kararlar vermesidir.**

Derin öğrenme, yapay zeka algoritmalarının yapısının katmanlı ve daha güçlü olduğu ve sinir ağı olarak adlandırılan özel bir makine öğrenimi kategorisidir. Derin öğrenme modelleri, mühendislerin doğruluğu artırmak ve modelin makine öğreniminin kapasitesinin ötesinde daha fazla nüans katmanını tanımasını sağlamak için modeli sürekli olarak ayarlayacağı tekrarlayan bir test döngüsü sürecinden geçirilmesidir.

Yapay Zeka Algoritmalarının Yaygın Türleri:

Yapay zeka modellerinin amacı, farklı türdeki birden çok girdi arasındaki ilişkiyi anlamaya çalışarak sonuçları tahmin etmek veya kararlar almak için bir veya daha fazla algoritma kullanmaktır. Yapay zeka modelleri, bu göreve nasıl yaklaştıklarına göre farklılık gösterir ve yapay zeka geliştiricileri, bir hedefe veya işleve ulaşmak için birden fazla algoritmayı birlikte devreye alabilir.

- Doğrusal regresyon, genellikle $f(t)=at+b$ şeklinde basit bir çizgi grafiğinde temsil edilen bir veya daha fazla X girişi(leri) ve Y çıkışı arasındaki doğrusal ilişkiyi eşleştirir.
- Lojistik regresyon, ikili bir X değişkeni (doğru veya yanlış, mevcut veya yok gibi) ile Y çıktısı arasındaki ilişkiyi eşleştirir.
- Doğrusal diskriminant analizi, başlangıç verilerinin ayrı kategoriler veya sınıflandırmalarla nitelendirilmesi dışında, lojistik regresyon gibi çalışır.
- Karar ağaçları, karar ağacı bir sonuca ulaşana kadar bir dizi giriş verisine dallanan mantık modelleri uygular.
- Naive Bayes, başlangıç verileri arasında hiçbir ilişki olmadığını varsayan bir sınıflandırma tekniğidir.
- K-en yakın komşu, benzer özelliklere sahip girdilerin korelasyonları grafiğe döküldüğünde (Öklid mesafesi açısından), bunların birbirine yakın olacağını varsayan bir sınıflandırma tekniğidir. Benzer özelliklere sahip girdi verileri arasında ilişkinin yönü ve gücü hesaplandığında bulunan değerlerin birbirine yakın olacağını varsayan bir sınıflandırma tekniğidir.
- Vektör nicelemeyi öğrenmek, k-en yakın komşuya benzer, ancak bireysel veri noktaları arasındaki mesafeyi ölçmek yerine model, veri noktaları gibi prototiplere yakınsar.

- Destek vektör makinesi algoritmaları, daha doğru sınıflandırma için veri noktalarını belirgin şekilde ayıran, hiper düzlem adı verilen bir bölücü oluşturur. Benzer özelliklere sahip girdi verilerini daha doğru sınıflandırmak için verilerin özelliklerini birbirlerinde belirgin şekilde ayıran, hiper düzlem adı verilen bir tampon bölgelerin sınırlarını oluşturur.
- Torbalama, daha doğru bir model oluşturmak için birden fazla algoritmayı birleştirirken, rastgele orman, daha doğru tahmin elde etmek için birden fazla karar ağacını bir araya getirir.
- Derin sinir ağları, girdilerin geçmesi gereken ve nihai bir tahmin veya karar noktası ile sonuçlanan birçok algoritma katmanından oluşan bir yapıyı ifade eder.

Yapay Zeka Modelleme için Teknoloji Gereksinimleri

Yapay zeka modelleri, o kadar büyüyor ki, onları verimli bir şekilde eğitmek için daha fazla veri gerekiyor ve ne kadar hızlı veri taşırsanız, modeli o kadar hızlı eğitebilir ve devreye alabilirsiniz. Bilgi işleme platformları, yüksek performanslı CPU'lar, yüksek kapasiteli depolama ve yoğun trafik akışını kaldırabilen yüksek bant genişliğine sahip ağ yapıları ile yapay zeka iş yükleri için ayarlanmış yapılandırmalar sağlamaya yardımcı olur.

Yapay Zeka / Makine Öğrenimi için Yazılım Çözümleri

Günümüzde piyasada bulunan çok sayıda makine öğrenimi ve derin öğrenme yazılımı seçeneği karşısında şaşkına dönmek kolaydır. Ancak Intel teklifleri ile tümü Intel® platformlarında performans için optimize edilmiş ortak çerçeveler ve kütüphaneler için tek bir kaynağa erişebilirsiniz.

- Sonuçlara hızlı zaman desteği sağlayan platformlar yapay zeka çıkarımını optimize edilmesine ve hızlandırılmasına olanak tanır. Bu platformlar, hem veri merkezi uygulamaları hem de uçta devreye almalarında yapay zeka destekli veri üretimi için faydalıdır.
- Önceden eğitilmiş yapay zeka modelleri, ilgili platformlarda performans için optimize edilmiş olan TensorFlow, PyTorch ve scikit-learn gibi ortak çerçevelerin dağıtımını içerir. Bu kaynaklar, geliştiricilerin yapay zeka modelleme çabalarını ve devreye alma süresini hızlandırmasına yardımcı olabilir.
- Derin öğrenme uygulamaları için tasarlanmış ve Apache Spark, TensorFlow, Keras ve BigDL çerçeveleri üzerine tasarlanmış birleşik bir yapay zeka ve analiz araçları platformu derin öğrenme projeleri için veri tabanı entegrasyonunu ve kolay başlatmayı kolaylaştırmak için yerleşik derin öğrenme modellerinin yanı sıra üst düzey soyutlamalar ve API'ler içerir.

Yapay Zeka analiz platformlarına entegre etmek için iş zekası ve yapay zeka alanında SAP, Oracle ve SAS kullanılmaktadır. Yapay zeka iş ortakları arasında Cloudera, Data Robot, Omni-Sci, H2O, SAS ve Splunk bulunmaktadır.

Makine Öğrenmesi

Makine Öğrenimi türleri

Makine Öğrenimi temel olarak üç türe ayrılabilir:

- **Denetimli Öğrenme:** Denetimli öğrenme, makinenin verilerden öğrenmek için harici denetime ihtiyaç duyduğu bir tür Makine öğrenimidir. **Denetimli öğrenme modelleri, etiketli veri kümesi kullanılarak eğitilir.** Regresyon ve Sınıflandırma, Denetimli Makine Öğrenimi modelleridir. Regresyon matematiksel ifadelerin katsayılarının belirlenmesine dayanır.
- **Denetimsiz Öğrenme:** Makinenin verilerden öğrenmek için herhangi bir dış denetime ihtiyaç duymadığı bir makine öğrenmesi türüdür, dolayısıyla denetimsiz öğrenme olarak adlandırılır. **Denetimsiz modeller, etiketlenmemiş veri kümesi kullanılarak eğitilebilir.** Bunlar, ilişkilendirme ve Kümeleme problemlerini çözmek için kullanılır.
- **Takviyeli Öğrenme:** Takviyeli öğrenmede, bir etmen, eylemler üreterek çevresiyle etkileşime girer ve geri bildirim yardımıyla öğrenir. Geri bildirim, temsilciye ödül şeklinde verilir, örneğin her iyi eylem için olumlu bir ödül alır ve her kötü eylem için olumsuz bir ödül alır. Temsilciye sağlanan herhangi bir denetim yoktur. Takviyeli öğrenmede Q-Learning algoritması kullanılır.

Regresyon:

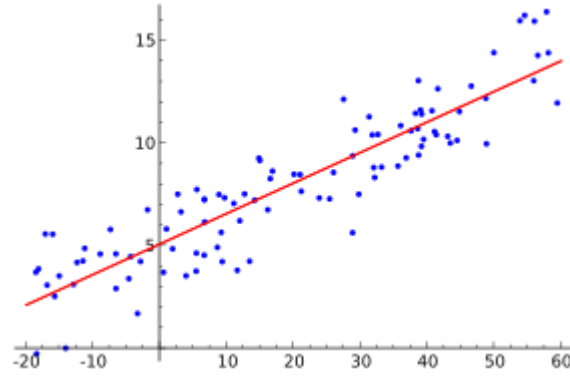
Sayısal değişkenler arasındaki ilişkiyi bulmaya çalıştığınızda regresyon yöntemi kullanılır. Makine Öğreniminde ve istatistiksel modellemede bu ilişki gelecekteki olayların sonucunu tahmin etmek için kullanılır. Eksik, hatalı ara değerleri belirlemede kullanılır. Sürekli değerleri tahmin etmek için regresyon algoritmaları kullanılır.

Regresyon algoritmaları, girdi özellikleri ile çıktı arasındaki ilişkiyi modelleyen istatistiksel bir yaklaşımdır. Regresyon algoritmasında veri kümesine ait kabul edilebilir birden fazla sayıda fonksiyon elde edilebilir. Bu durumda doğru fonksiyonu seçimde ortalama karelerin hatası göz önüne alınır.

Regresyon algoritmaları:

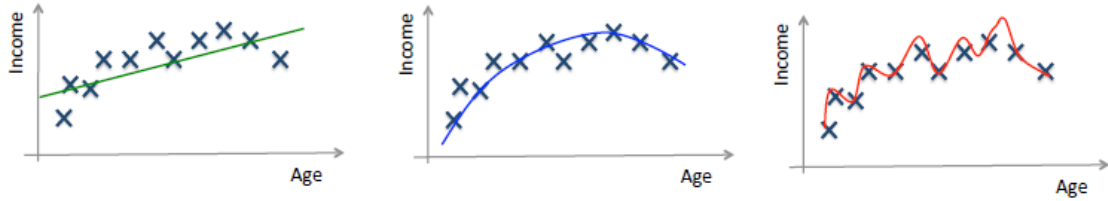
- Linear Regression
- Polynomial Regression
- Exponential Regression
- Logistic Regression
- Logarithmic Regression

Regresyon analizi, girdi değişkenleri ile çıktı arasındaki ilişkiyi tahmin etmek için çok çeşitli istatistiksel yöntemleri kapsar. En yaygın şekli, sıradan en küçük kareler gibi bir matematiksel kritere göre verilen verilere en iyi uyacak şekilde tek bir çizginin çizildiği doğrusal regresyondur.



Makine öğrenmesinde bir modeli eğitmenin amacı, tüm örneklerde ortalama olarak düşük kayıplı bir eşik değeri bulmaktır. Bu ortalama kareler hatası ile bulunur; farkların kareleri toplamının örnek sayısına bölümüdür.

Regresyon katsayıları bulunurken optimal kapsayıcı bir eğri fonksiyonu elde edilmelidir.
İpucu: Sınıflandırma yapılırken yüksek dereceli polinomlardan kaçınılmalıdır.



$$f(x) = \lambda_0 + \lambda_1 x \dots (1)$$

$$f(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2 \dots (2)$$

$$f(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3 + \lambda_4 x^4 \dots (3)$$

Yeterli veri olduğunda, aşırı uyum sorununu önlemek için "İzotonik Regresyon" kullanılır.

Doğrusal Regrasyon

Doğrusal Regresyon Nedir? Makine Öğreniminde Nasıl Kullanılır?

Çoğu makine öğrenimi (ML) algoritması, başta istatistik olmak üzere çeşitli alanlardan ödünç alınmıştır. Modellerin daha iyi tahmin etmesine yardımcı olabilen her şey, sonunda makine öğreniminin bir parçası haline gelmektedir. Bu nedenle, doğrusal regresyonun hem istatistiksel hem de makine öğrenimi algoritması olduğunu söylemek güvenlidir.

Doğrusal regresyon, veri bilimi ve makine öğreniminde kullanılan popüler ve karmaşık olmayan bir algoritmadır. Denetimli bir öğrenme algoritmasıdır ve değişkenler arasındaki matematiksel ilişkiyi incelemek için kullanılan en basit regresyon biçimidir.

Örneğin,

Giriş ile çıkış arasındaki ifade, $f(t)=at+b$ olsun. Bağımsız değişken: t , Bağımlı değişken: $f(t)$ olur.

Veri yığınının a ve b katsayıları bulunması istenir. Veri yığınının zaman zaman aralığı, örneğin 1 saniye 120 saniye arasında giriş ve çıkış değerleri var. 60 ile 65 saniye arasındaki değerler eksik. 60 ile 65 saniye arasındaki değerleri 0 alarak ya da optimum değerler vererek $f(t)$ çıkış fonksiyonunun a ve b katsayıları belirlenmiş olur. Eksik değerler tamamlandı. 125 saniyede ne olacak diyer sorduğumda $t=125$, girildiğinde $f(t)$ değeri bulunmuş olur. Bu değer tahmindir.

Doğrusal regresyon, değişkenler arasındaki ilişkiyi göstermeye çalışan istatistiksel bir yöntemdir. Farklı veri noktalarına bakar ve bir trend çizgisi çizer. Daha kesin olarak, bir bağımlı değişken ile bir dizi diğer bağımsız değişken arasındaki ilişkinin karakterini ve gücünü belirlemek için doğrusal regresyon kullanılır. Bir şirketin hisse senedi fiyatını tahmin etmek gibi tahminler yapmak için modeller oluşturmaya yardımcı olur. **Gözlemlenen veri kümesine bir regresyon modeli uydurmaya çalışmadan önce, bağımlı ve bağımsız değişkenler arasında bir ilişki olup olmadığı belirlenmelidir.** Aralarında gözle görülür bir korelasyon olmalıdır.

Örneğin, daha yüksek üniversite notları mutlaka daha yüksek bir maaş paketi anlamına gelmez. İki değişken arasında bir ilişkinin belirlenmesinde dağılım grafiği oluşturmak yani bağımsız ve bağımlı değişkenler arasındaki ilişkinin gücünü belirlemek için idealdir. Dağılım grafiği herhangi bir artan veya azalan eğilim göstermiyorsa, gözlemlenen değerlere doğrusal bir regresyon modeli uygulamak faydalı olmayabilir.

Korelasyon katsayıları, iki değişken arasındaki ilişkinin ne kadar güçlü olduğunu hesaplamak için kullanılır. Genellikle r ile gösterilir ve -1 ile 1 arasında bir değere sahiptir. Pozitif bir korelasyon katsayısı değeri, değişkenler arasında pozitif bir ilişkiyi gösterir. Aynı şekilde,

negatif bir deęer, deęişkenler arasında negatif bir ilişkiyi gösterir. Eęer r deęeri sıfıra yakın deęerler alıyor ise iki deęişken arasında bir ilişki olmadığı sonucuna varılır.

$$y = mx + b$$

Bu basit doğrusal regresyon denkleminde:

- y tahmini bağımlı deęişkendir (veya çıktıdır)
- m, regresyon katsayısıdır (veya eğimdir)
- x, bağımsız deęişkendir (veya girdidir)
- b sabittir (veya y-keseni üzerinde bir nokta)

Pearson Çarpım Moment Korelasyon Katsayısı:

$$r = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{n(\sum x_i^2) - (\sum x_i)^2} \sqrt{n(\sum y_i^2) - (\sum y_i)^2}}$$

Korelasyon katsayısı abs(r)'nin yorumu;

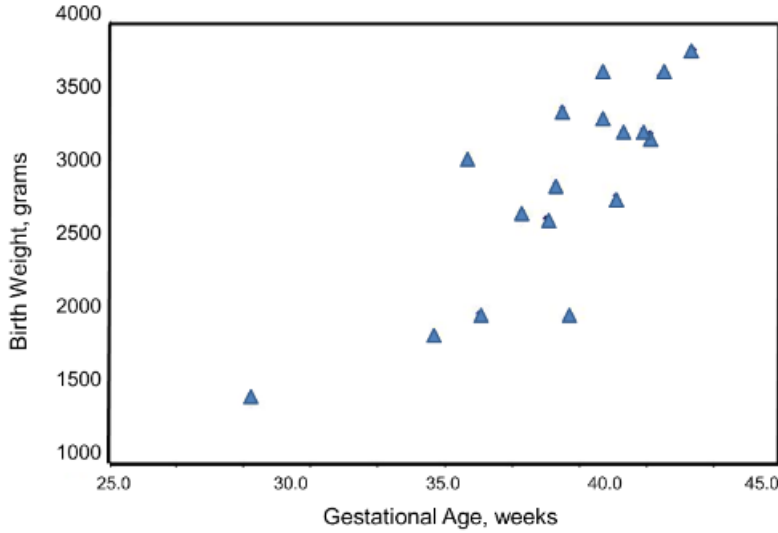
- $r < 0.2$ ise çok zayıf ilişki yada korelasyon yok
- 0.2-0.4 arasında ise zayıf korelasyon
- 0.4-0.6 arasında ise orta şiddette korelasyon
- 0.6-0.8 arasında ise yüksek korelasyon
- $0.8 >$ ise çok yüksek korelasyon olduğu yorumu yapılır.

Örnek:

Hafta olarak ölçülen doğumdaki gebelik yaşı ile gram cinsinden ölçülen doğum ağırlığı arasındaki ilişkiyi araştırmak için 17 bebeęi içeren küçük bir çalışma yapılmıştır.

Infant ID #	Gestational Age (weeks)	Birth Weight (grams)
1	34.7	1895
2	36.0	2030
3	29.3	1440
4	40.1	2835
5	35.7	3090
6	42.4	3827
7	40.3	3260
8	37.3	2690
9	40.9	3285
10	38.3	2920
11	38.5	3430
12	41.4	3657
13	39.7	3685
14	39.7	3345
15	41.1	3260
16	38.0	2680
17	38.7	2005

Gebelik yaşı ile bebek doğum ağırlığı arasındaki ilişkiyi tahmin etmek istiyoruz. Bu örnekte, doğum ağırlığı bağımlı değişken ve gebelik yaşı bağımsız değişkendir. Böylece y =doğum ağırlığı ve x =gebelik yaşı. Veriler, aşağıdaki şekilde bir dağılım diyagramında gösterilmektedir.



Her nokta bir (x,y) çiftini temsil eder (bu durumda hafta cinsinden ölçülen gebelik yaşı ve gram cinsinden ölçülen doğum ağırlığı). Bağımsız değişken olan gebelik yaşının yatay ekseninde (veya X ekseninde) ve bağımlı değişkenin (doğum ağırlığı) dikey ekseninde (veya Y ekseninde) olduğuna dikkat edin. Dağılım grafiği, gebelik yaşı ile doğum ağırlığı arasında pozitif veya doğrudan bir ilişki gösterir. Gebelik yaşı daha kısa olan bebeklerin daha düşük ağırlıkla doğma olasılığı daha yüksektir ve gebelik yaşı daha uzun olan bebeklerin daha yüksek ağırlıkla doğma olasılığı daha yüksektir.

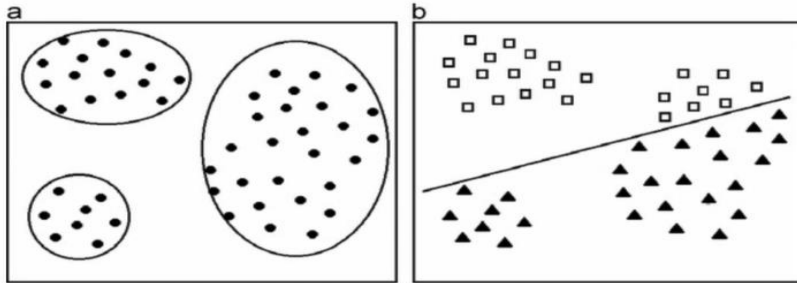
Korelasyon katsayısı, $r=0.82$ hesaplanır.

Kümeleme (Clustering)

Verilerin yakınlık, uzaklık, benzerlik gibi ölçütlere göre analiz edilerek sınıflara ayrılmasına kümeleme denir. Kümeleme, denetimsiz öğrenmenin bir yöntemidir ve birçok alanda kullanılan istatistiksel veri analizi için yaygın bir tekniktir. Denetimsiz öğrenme, veri kümesi ile çıktıların olmadığı bir öğrenme metodudur. Veri kümesindeki verileri yorumlayarak ortak noktaları bulmak ve bunları kümeleştirme işlemi yapılarak anlamlı bir veri elde edebilmektir. Sistem, öğreten olmadan öğrenmeye çalışır. Ham verileri organize verilere dönüştüren bir makine öğrenimi türüdür.

Denetimsiz öğrenmede sadece sınıflandırılmamış veriler vardır bu verilerden sonuçlar çıkarılmaya çalışılır. Veriyi değişkenler arasındaki ilişkilere dayalı olarak kümeleyerek çeşitli modeller, yapılar oluşturulur.

Örneğin, bir alışveriş sitesinde alınan bir ürünün yanında kullanıcıların alabileceği diğer ürünlerin tavsiye olarak belirlenmesi. Ya da bir hizmet satın alındığında, o hizmetle etkileşimli diğer hizmetlerin müşterinin ilgi alanına girmesi.



Kümeleme

Sınıflandırma

Kümeleme algoritmaları, veri kümesindeki bileşenleri kendi aralarında benzerliklerinden gruplamaya çalışır. Burada kaç grup olacağı veya en uygun küme sayısını algoritmanın kendisi belirler.

Örnek: Bir sandık dolusu meyve içerisinde bir tane elmayı belirlesin. Ardından tüm elmaları bulan bir senaryo sizce nasıl olmalıdır? En sonunda sandıkta bir tane elma kalmaması yazdığın senaryonun hangi özelliği sağlanmış olur. Performans artmış, deneyim kazanmış olur.

Alışveriş yapılan bir markette kasiyerin bir robot olduğun düşünün ve tüm ürünler birbirine karışmış olsun. Elma bulup, onu tanıyıp diğer elemanları yığın içerisinde topladığını düşünün. Seçme işlemi devam ederken yetenek kazanarak performansını atırabilir; hatalı seçtiği ürünler var ise ayıklayabilir. Böylece ürünlerin birbirlerine benzeme yakınlığı uzaklaşarak, ayırım yapma yeteneği artırılmış olur. Böylece sınıflandırma da yapılmış olur.

Elmalar da kendi aralarında kümeleme yapılabilir mi? Aynı tipte verilerin değişik segmentlere bölünmüş halidir. Elinde örnekler var ama hangi veya kaç sınıfa ait olduğunu bilmiyor. Sınıfları(kümeleri) kendisi inşaa ediyor. Örneğin elimizde sadece domatesler varsa, ve bunları kalitelerine göre ayrılırsa bu kümeleme işlemidir.

Kümelemenin uygulama alanları:

Tıp'da elde edilen görüntülemeler üzerindeki farklıları analiz edilerek değişik nitelikler çıkartılabilir.

Suç Yerlerinin Belirlenmesi: Bir şehirdeki belirli bölgelerde mevcut olan suçlarla ilgili veriler, suç kategorisi, suç alanı ve ikisi arasındaki ilişki, bir şehirdeki ya da bölgedeki suça eğilimli alanlara ilişkin kaliteli bilgiler verebilir.

Oyuncu istatistiklerini analiz etmek: Oyuncu istatistiklerini analiz etmek, spor dünyasının her zaman kritik bir unsuru olmuştur ve artan rekabetle birlikte, makine öğrenmenin burada oynayacağı kritik bir rol vardır.

Çağrı Kaydı Detay Analizi: Bir çağrı detay kaydı (CDR), telekom şirketleri tarafından bir müşterinin araması, SMS ve internet etkinliği sırasında elde edilen bilgilerdir. Bu bilgiler, müşteri demografisiyle birlikte kullanıldığında, müşterinin ihtiyaçları hakkında daha fazla bilgi sağlar.

Müşteri Segmentasyonu:

Collaborative Filtering: Davranışları birbirine benzeyen insanların yaptıklarına bakılarak kümedeki birinin ne yapacağını kestirmek.

Tehdit ve Sahtekarlık Yakalama: Sınıflandırmada tehditlerin özellikleri makineye öğretilir. Kümelemede ise kümelerin dışında kalan , herhangi bir kümeye girmeyen örnek bir tehdit unsuru olarak tanımlanabilir.

Eksik Verilerin Tanımlanması: Örneğin birinin maaş bilgisi eksikse segmentte benzer kişilerin maaş ortalamasına göre bir maaş hesaplanabilir.

Pazar Segmentasyonu:

Davranışsal Segmentasyon: Örneğin ücretsiz uygulamaları yükleyenler neyi seçiyorlar?

Demografik Segmentasyon: Müşterilerin yaşı, cinsiyeti, eğitim düzeyi ile davranışlarının entegre edilmesi.

Psikolojik Segmentasyon: Müşterilerin hayal-beklentilerine göre farklı ürünler sunulabilir.

Coğrafi Segmentasyon: Ülkelere şehirlere göre vs.

Kümeleme Çeşitleri:

- Hiyerarşik Kümeleme
- Gürültülü Uygulamaların Yoğunluğa Dayalı Konumsal Kümelenmesi (DBSCAN)
- K-means Kümeleme
- Ağırlık Ortalama Kaydırma Kümelemesi
- Gauss Karışım Modelleri (GMM) kullanarak Beklenti-Maksimizasyon (EM) Kümeleme

Ağırlık Ortalama Kaydırma Kümelemesi:

Ortalama kaydırma kümeleme, veri noktalarının yoğun alanlarını bulmaya çalışan kayan pencere tabanlı bir algoritmadır. Centroid tabanlı bir algoritmadır, yani amacın her bir grubun / sınıfın merkez noktalarını bulmaktır, bu da kayan pencere içindeki noktaların

ortalaması olacak merkez noktaları için adayları güncelleyerek çalışır. Bu aday pencereler daha sonra, neredeyse kopyaları ortadan kaldırmak için bir işlem sonrası aşamasında filtrelendir ve nihai merkez noktaları ve bunlara karşılık gelen grupları oluşturur.

Sürgülü pencerelerin tümü ile uçtan uca tüm sürecin bir örneği aşağıda gösterilmiştir. Her siyah nokta, kayan bir pencerenin merkezini temsil eder ve her gri nokta bir veri noktasıdır.

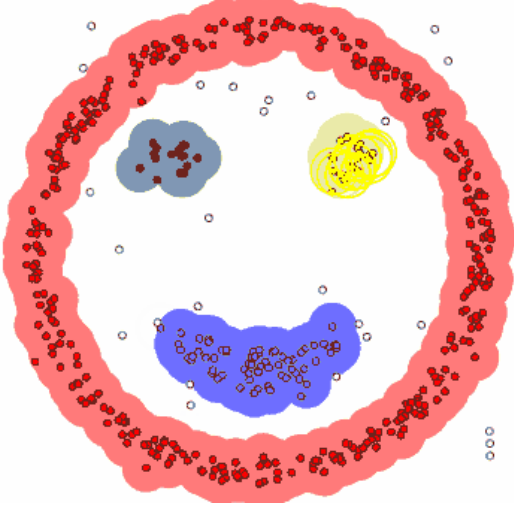
Sınav sorusu: En son ağırlık ortalama burasıdır denmesi için nasıl bir senaryo hazırlanmalı?



- 1) Ortalama-kaymayı açıklamak için, yukarıdaki resimde olduğu gibi iki boyutlu uzayda bir dizi noktayı ele alacağız. Bir C noktasında ortalanmış (rastgele seçilmiş) ve çekirdek olarak r yarıçapına sahip dairesel bir kayan pencere ile başlıyoruz. Ortalama kayma, bu çekirdeği yakınsamaya kadar her adımda yinelemeli olarak daha yüksek yoğunluklu bir bölgeye kaydırmayı içeren bir tepe tırmanma algoritmasıdır.
- 2) Her yinelemede, kayan pencere, merkez noktası pencere içindeki noktaların ortalamasına kaydırılarak (dolayısıyla adı) daha yüksek yoğunluklu bölgelere kaydırılır. Sürgülü pencere içindeki yoğunluk, içindeki noktaların sayısı ile orantılıdır. Doğal olarak, penceredeki noktaların ortalamasına geçerek, yavaş yavaş daha yüksek nokta yoğunluğuna sahip alanlara doğru hareket edecektir.
- 3) Bir kaymanın çekirdek içinde daha fazla noktayı barındırabileceği bir yön olmayana kadar ortalama göre kayan pencereyi kaydırmaya devam ediyoruz. Yukarıdaki grafiğe bakın; Artık yoğunluğu artırmayana kadar (yani penceredeki nokta sayısı) daireyi hareket ettirmeye devam ediyoruz.
- 4) 1'den 3'e kadar olan bu adım süreci, tüm noktalar bir pencerenin içinde kalana kadar birçok sürgülü pencerede yapılır. Birden çok sürgülü pencere örtüştüğünde, en çok noktayı içeren pencere korunur. Veri noktaları daha sonra bulundukları kayan pencereye göre kümelenir.

Gürültülü Uygulamaların Yoğunluğa Dayalı Konumsal Kümelenmesi (DBSCAN):

DBSCAN, ortalama kaymaya ekseninde, benzer ve yoğunluklu bölgeleri kümeleyen bir algoritmadır, ancak birkaç önemli avantajı vardır. Minimum bölge sayısında ve uzaklıkta maksimum yoğunluk bölgesi oluşturulması hedeflenir.

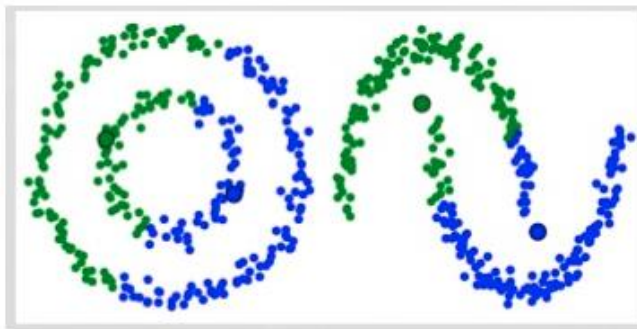


- DBSCAN, keyfi bir başlangıç veri noktasıyla başlar. Bu noktanın komşuluğu bir epsilon ϵ uzaklığı kullanılarak çıkarılır (ϵ mesafesi içindeki tüm noktalar komşuluk noktalarıdır).
- Bu mahalle içinde yeterli sayıda nokta (minPoints'e göre) varsa, kümeleme işlemi başlar ve mevcut veri noktası yeni kümedeki ilk nokta olur. Aksi takdirde, nokta gürültü olarak etiketlenecektir (daha sonra bu gürültülü nokta kümenin parçası haline gelebilir). Her iki durumda da bu nokta "ziyaret edildi" olarak işaretlenir.
- Yeni kümedeki bu ilk nokta için, ϵ mesafesi komşuluğundaki noktalar da aynı kümenin parçası olur. ϵ mahallesindeki tüm noktaları aynı kümeye ait yapma prosedürü, küme grubuna yeni eklenen tüm yeni noktalar için tekrarlanır.
- 2. ve 3. adımlardan oluşan bu süreç, kümedeki tüm noktalar belirlenene kadar, yani kümenin ϵ mahallesindeki tüm noktalar ziyaret edilip etiketlenene kadar tekrar edilir. Mevcut kümeyle işimiz bittiğinde, yeni bir ziyaret edilmemiş nokta alınır ve işlenir, bu da başka bir küme veya gürültü keşfine yol açar. Bu işlem, tüm noktalar ziyaret edildi olarak işaretlenene kadar tekrar eder.
- Bunun sonunda tüm noktalar ziyaret edildiğinden, her nokta bir kümeye ait veya gürültü olarak işaretlenecektir.

DBSCAN, diğer kümeleme algoritmalarına göre bazı büyük avantajlar sunar. İlk olarak, hiç bir küme gerektirmez. Ayrıca, veri noktası çok farklı olsa bile, aykırı değerleri basitçe bir kümeye atan ortalama kaymanın aksine, gürültü olarak tanımlar. Ek olarak, keyfi olarak boyutlandırılmış ve keyfi olarak şekillendirilmiş kümeleri oldukça iyi bulabilir. DBSCAN'ın ana dezavantajı, kümeler farklı yoğunlukta olduğunda diğerleri kadar iyi performans göstermemesidir. Bunun nedeni, komşu noktaların belirlenmesi için mesafe eşiği ϵ ve minPoints'in, yoğunluk değiştiğinde kümeden kümeye değişmesidir. Bu dezavantaj, çok yüksek boyutlu verilerde de ortaya çıkar, çünkü yine mesafe eşiğini ϵ tahmin etmek zorlaşır.

Gauss Karışım Modelleri (GMM) kullanarak Beklenti-Maksimizasyon (EM) Kümeleme:

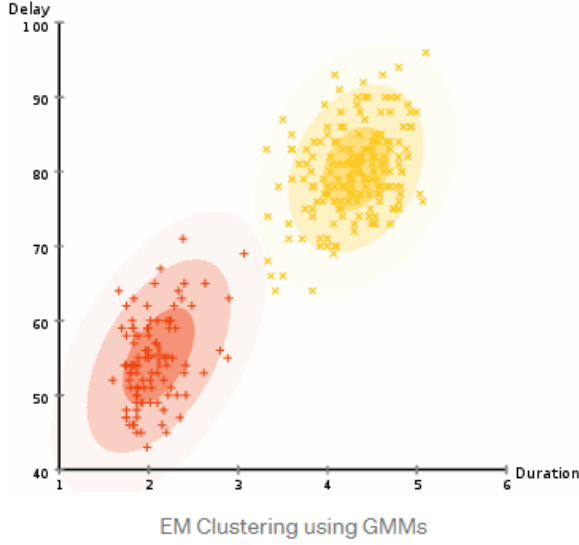
K-Ortalamlarının en büyük dezavantajlarından biri, küme merkezi için ortalama değerin naif kullanımıdır. Aşağıdaki resme bakarak bunun bir şeyleri yapmanın en iyi yolu olmadığını anlayabiliriz. Sol tarafta, aynı ortalamaya merkezlenmiş farklı yarıçaplara sahip iki dairesel küme olduğu insan gözüne oldukça açık görünüyor. K-Ortalamlar bunun üstesinden gelemez çünkü kümelerin ortalama değerleri birbirine çok yakındır. K-Ortalamlar, yine ortalamanın küme merkezi olarak kullanılması sonucunda kümelerin dairesel olmadığı durumlarda başarısız olur.



Two failure cases for K-Means

Gauss Karışım Modelleri (GMM'ler) bize K-Ortalamlarından daha fazla esneklik sağlar. GMM'ler ile veri noktalarının Gauss olarak dağıtıldığını varsayıyoruz; bu, ortalamayı kullanarak döngüsel olduklarını söylemekten daha az kısıtlayıcı bir varsayımdır. Bu şekilde, kümelerin şeklini açıklamak için iki parametremiz var: ortalama ve standart sapma! İki boyutlu bir örnek alırsak, bu, kümelerin her türlü eliptik şekli alabileceği anlamına gelir (çünkü hem x hem de y yönlerinde standart bir sapmaya sahibiz). Böylece, her Gauss dağılımı tek bir kümeye atanır.

Her küme için Gauss'un parametrelerini bulmak için (örneğin ortalama ve standart sapma), Beklenti-Maksimizasyon (EM) adı verilen bir optimizasyon algoritması kullanacağız. Kümelere uydurulan Gaussian'ların bir örneği olarak aşağıdaki grafiğe bir göz atın. Daha sonra GMM'leri kullanarak Beklenti-Maksimizasyon kümeleme sürecine geçebiliriz.



- 1) Küme sayısını seçerek (K-Means'ın yaptığı gibi) ve her küme için Gauss dağılım parametrelerini rastgele başlatarak başlıyoruz. Verilere de hızlıca göz atarak ilk parametreler için iyi bir tahmin sağlamaya çalışılabilir. Yine de, yukarıdaki grafikte de görülebileceği gibi, bu% 100 gerekli değildir çünkü Gauss bize çok zayıf olarak başlarlar, ancak hızla optimize edilirler.
- 2) Her küme için bu Gauss dağılımları göz önüne alındığında, her veri noktasının belirli bir kümeye ait olma olasılığı hesaplanır. Bir nokta Gauss'un merkezine ne kadar yakınsa, o kümeye ait olma olasılığı o kadar artar. Gauss dağılımında verilerin çoğunun kümenin merkezine daha yakın olduğunu varsaydığımız için, bu sezgisel bir anlam ifade etmelidir.
- 3) Bu olasılıklara dayanarak, kümeler içindeki veri noktalarının olasılıklarını maksimize edecek şekilde Gauss dağılımları için yeni bir dizi parametre hesaplıyoruz. Bu yeni parametreleri, veri noktası konumlarının ağırlıklı toplamını kullanarak hesaplıyoruz; burada ağırlıklar, o belirli kümeye ait veri noktasının olasılıklarıdır. Bunu görsel olarak açıklamak için yukarıdaki grafiğe, özellikle örnek olarak sarı kümeye bakabiliriz. Dağıtım ilk yinelemede rastgele başlar, ancak sarı noktaların çoğunun bu dağılımın sağında olduğunu görebiliriz. Olasılıklara göre ağırlıklandırılmış bir toplamı hesapladığımızda, merkeze yakın bazı noktalar olmasına rağmen çoğu sağdadır. Dolayısıyla, doğal olarak dağılımın ortalaması bu noktalar kümesine daha da yaklaşır. Ayrıca noktaların çoğunun “üstten-sağdan-aşağıya” olduğunu da görebiliriz. Bu nedenle standart sapma, olasılıkların ağırlıklandığı toplamı maksimize etmek için bu noktalara daha uygun bir elips oluşturmak üzere değişir.
- 4) Dağıtımların yinelemeden yinelemeye pek değişmediği yakınsamaya kadar 2. ve 3. adımlar yinelemeli olarak tekrarlanır.

GMM kullanmanın 2 önemli avantajı vardır. Birincisi, GMM'ler küme kovaryansı açısından K-Ortalamalarına göre çok daha esnektir; standart sapma parametresi nedeniyle, kümeler dairelerle sınırlı olmak yerine herhangi bir elips şeklini alabilir. K-Ortalamaları aslında her kümenin tüm boyutlardaki kovaryansının 0'a yaklaştığı özel bir GMM durumudur. İkinci olarak, GMM'ler olasılıkları kullandığından, veri noktası başına birden çok kümeye sahip olabilirler. Dolayısıyla, bir veri noktası üst üste binen iki kümenin ortasındaysa, sınıf 1'e yüzde X ve yüzde Y yüzde 2'ye ait olduğunu söyleyerek sınıfını tanımlayabiliriz. Yani GMM'ler karma üyeliği destekler.

Kümeleme Türleri:

Kümeleme, veri noktalarının benzerlik derecelerine göre farklı kümeler halinde gruplandırıldığı bir denetimsiz öğrenme türüdür.

Çeşitli kümeleme türleri şunlardır:

- Hiyerarşik kümeleme
- Bölümleme kümeleme

Hiyerarşik kümeleme ayrıca alt bölümlere ayrılır:

- Toplu kümeleme
- Bölücü kümeleme

Bölümleme kümeleme ayrıca şu alt bölümlere ayrılır:

- K-Ortalamalar kümeleme
- Bulanık C-Ortalamalar kümeleme

Lineer regresyonda anahtar terminolojiler

Doğrusal regresyon analizini anlamak, bir dizi yeni terime aşina olmak anlamına da gelebilir. İstatistik veya makine öğrenimi dünyasına yeni adım attıysanız, bu terminolojileri adil bir şekilde anlamanız önemli olacaktır.

- **Değişken (Variable):** Sayılabilen veya ölçülebilen herhangi bir sayı, nicelik veya özelliktir. Veri ögesi olarak da adlandırılır. Gelir, yaş, hız ve cinsiyet örnek olarak verilebilir.
- **Katsayı (Coefficient):** Yanındaki değişkenle çarpılan bir sayıdır (genellikle bir tam sayıdır). Örneğin, $7x$ 'te 7 sayısı katsayıdır. $y = a x + b$, a ve b katsayıdır, x : bağımsız değişken, y ise bağımlı değişkendir.
- **Aykırı Değerler (Outliers):** Bunlar, diğerlerinden önemli ölçüde farklı olan veri noktalarıdır.
- **Kovaryans (Covariance):** İki değişken arasındaki doğrusal ilişkinin yönü. Başka bir deyişle, iki değişkenin doğrusal olarak ilişkili olma derecesini hesaplar.
- **Çok değişkenli (Multivariate):** Tek bir sonuçla sonuçlanan iki veya daha fazla bağımsız değişkenlerin dahil edilmesi anlamına gelir.
- **Artıklar (Residuals):** Bağımlı değişkenin gözlemlenen ve tahmin edilen değerleri arasındaki fark.
- **Değişkenlik (Variability):** Tutarlılığın olmaması veya bir dağılımın ne ölçüde sıkıştırıldığı veya esnetildiği.
- **Doğrusallık (Linearity):** Orantılılıkla yakından ilgili olan ve grafiksel olarak düz bir çizgi olarak gösterilebilen bir matematiksel ilişkinin özelliği.
- **Doğrusal fonksiyon (Linear function):** Grafiği düz bir çizgi olan fonksiyondur.
- **Doğrusallık (Collinearity):** Bir regresyon modelinde doğrusal bir ilişki sergileyecek şekilde bağımsız değişkenler arasındaki korelasyon.
- **Standart sapma (Standard deviation - SD):** Bir veri kümesinin ortalamasına göre dağılımının bir ölçüsüdür. Başka bir deyişle, sayıların ne kadar dağıldığının bir ölçüsüdür.
- **Standart hata (Standard error - SE):** İstatistiksel bir örneklem popülasyonunun yaklaşık SD'si. Değişkenliği ölçmek için kullanılır.

ML modelinin performansını deęerlendirmesi

ML modelinin performansını deęerlendirmenin bazı popöler yolları şunlardır:

- **Karıřıklık Matrisi:** Makine öęrenmesinde sınıflandırma modelinin performansını belirlemek için kullanılan farklı deęer setlerine sahip $N \times N$ tablosudur.
- **F1 puanı:** ML modelini deęerlendirmek için en iyi ölçütlerden biri olarak kullanılan kesinlik ve hatırlamanın harmonik ortalamasıdır.
- **Kazanç ve artış çizelgeleri:** Kazanç ve Yükseliş çizelgeleri, olasılıkların sıra sırasını belirlemek için kullanılır.
- **AUC-ROC eęrisi:** AUC-ROC, başka bir performans ölçüsüdür. ROC, hassasiyet arasındaki alandır. Area Under the Curve (AUC) of Receiver Characteristic Operator (ROC). AUC-ROC eęrisinin makine öęrenimi sınıflandırının ne kadar iyi performans gösterdiğini görselleştirilmesine yardımcı olur. Sadece ikili sınıflandırma problemlerinde işe yaramasına rağmen, çok sınıflı sınıflandırma problemlerini de deęerlendirmek için genişletebilmektedir. Deęerlendirme kriterleri - AUC ne kadar yüksekse model o kadar iyidir.

Karıřıklık matrisi (confusion matrix):

		ACTUAL VALUES	
		POSITIVE	NEGATIVE
PREDICTED VALUES	POSITIVE	TP	FP
	NEGATIVE	FN	TN

Duyarlılık / Gerçek Pozitif Oranı (TPR - True Positive Rate) / Geri Çağırma:

Duyarlılık (Sensitivity) pozitif sınıfın ne kadarının doğru sınıflandırıldığını söyler. Basit bir örnek, gerçek hasta insanların ne kadarının model tarafından doğru bir şekilde tespit edildiğini belirlemek olabilir.

$$Duyarlılık = \frac{TP}{TP + FN}$$

Yanlış Negatif Oran (FNR - False Negative Rate), pozitif sınıfın ne kadarının yanlış sınıflandırıldığını söyler. Pozitif sınıfı doğru bir şekilde sınıflandırmak istediğimiz için daha yüksek bir TPR ve daha düşük bir FNR arzu edilir.

$$FNR = \frac{FN}{TP + FN}$$

Özgüllük / Gerçek Negatif Oran (True Negative Rate):

Özgüllük negatif sınıfın ne kadarının doğru sınıflandırıldığını söyler. Duyarlılık ile aynı örneği alırsak, Özgüllük, model tarafından doğru bir şekilde tanımlanmış sağlıklı insanların oranını belirlemek anlamına gelir.

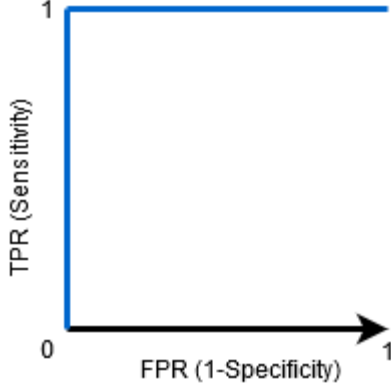
$$Özgüllük = \frac{TN}{TN + FP}$$

Yanlış Pozitif Oran (FPR - False Positive Rate):

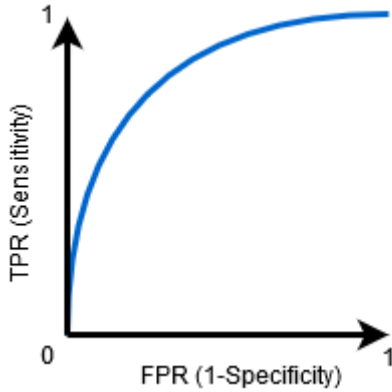
FPR, sınıflandırıcı tarafından negatif sınıfın ne kadarının yanlış sınıflandırıldığını söyler. Negatif sınıfı doğru bir şekilde sınıflandırmak istediğimiz için daha yüksek bir TNR ve daha düşük bir FPR arzu edilir. Bu ölçütlerden Duyarlılık ve Özgüllük belki de en önemlileridir ve bunların bir değerlendirme ölçütü oluşturmak için nasıl kullanıldığını daha sonra göreceğiz. Ancak ondan önce, tahmin olasılığının neden hedef sınıfı doğrudan tahmin etmekten daha iyi olduğunu anlayalım.

$$FPR = \frac{FP}{TN + FP} = 1 - Duyarlılık$$

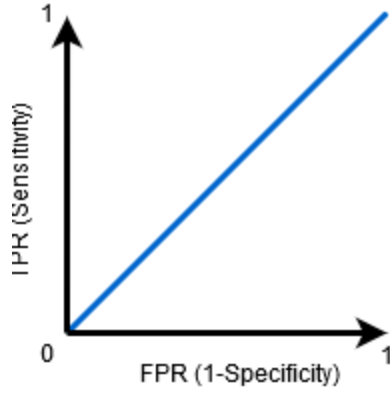
Alıcı Operatör Karakteristik (ROC) eğrisi, ikili sınıflandırma sorunları için bir değerlendirme metriğidir. Çeşitli eşik değerlerinde FPR'ye karşı TPR'yi çizen ve esas olarak "sinyali" "gürültüden" ayıran bir olasılık eğrisidir. Eğrinin Altındaki Alan (AUC), bir sınıflandırıcının sınıfları ayırt etme yeteneğinin ölçüsüdür ve ROC eğrisinin bir özeti olarak kullanılır. AUC ne kadar yüksek olursa, modelin pozitif ve negatif sınıfları ayırt etmedeki performansı o kadar iyi olur.



AUC = 1 olduğunda, sınıflandırıcı tüm Pozitif ve Negatif sınıf puanlarını doğru bir şekilde mükemmel bir şekilde ayırt edebilir. Bununla birlikte, AUC 0 olsaydı, sınıflandırıcı tüm Negatifleri Pozitif olarak ve tüm Pozitifleri Negatif olarak tahmin ediyor olurdu.



$0.5 < AUC < 1$ olduğunda, sınıflandırıcının pozitif sınıf değerlerini negatif sınıf değerlerinden ayırt edebilme olasılığı yüksektir. Bunun nedeni, sınıflandırıcının Yanlış negatifler ve Yanlış pozitiflerden daha fazla sayıda Doğru pozitif ve True negatif tespit edebilmesidir.



AUC=0.5 olduğunda, sınıflandırıcı Pozitif ve Negatif sınıf puanları arasında ayrım yapamaz. Yani, sınıflandırıcı tüm veri noktaları için rastgele sınıfı veya sabit sınıfı tahmin ediyor. Dolayısıyla, bir sınıflandırıcının AUC değeri ne kadar yüksek olursa, pozitif ve negatif sınıfları ayırt etme yeteneği o kadar iyi olur.

Örnek:

Point C			Point D		
Actual			Actual		
Predicted	+ve	-ve	Predicted	+ve	-ve
	4	2		3	2
-ve	1	3	-ve	2	3

TPR/Sensitivity = 0.8	TPR/Sensitivity = 0.6
FPR = 0.4	FPR = 0.4
Specificity = 1-FPR = 0.6	Specificity = 1-FPR = 0.6

C ve D noktaları arasında, aynı Özgüllük için C noktasındaki Hassasiyet D noktasından daha yüksektir. Bu, aynı sayıda yanlış sınıflandırılmış Negatif sınıf puanı için sınıflandırıcının daha yüksek sayıda Pozitif sınıf puanı öngördüğü anlamına gelir. Bu nedenle, C noktasındaki eşik D noktasından daha iyidir.

E Noktası, Özgüllüğün en yüksek olduğu yerdir. Yani model tarafından sınıflandırılan Yanlış Pozitifler yoktur. Model, tüm Negatif sınıf puanlarını doğru bir şekilde sınıflandırabilir! Sorunumuz, kullanıcılarımıza mükemmel şarkı önerileri vermek olsaydı, bu noktayı seçerdik.

		Point E	
		Actual	
		+ve	-ve
Predicted	+ve	3	0
	-ve	2	5

TPR/Sensitivity = 0.6
 FPR = 0
 Specificity = 1-FPR = 1

Bu mantıkla hareket ederek, mükemmel bir sınıflandırıcıya karşılık gelen noktanın grafikte nerede olduğunu tahmin edebilir misiniz? Evet! Kartezyen düzlemde (0, 1) koordinatına karşılık gelen ROC grafiğinin sol üst köşesinde olacaktır. Burada hem Duyarlılık hem de Özgüllük en yüksek olur ve sınıflandırıcı tüm Pozitif ve Negatif sınıf noktalarını doğru bir şekilde sınıflandırır.

- **Gini Katsayısı:** Gini İndeksi olarak da bilinen sınıflandırma problemlerinde kullanılır. Değişkenlerin değerleri arasındaki belirsizliği belirler. Gini'nin yüksek değeri iyi bir modeli temsil eder.
- **Kök ortalama kare hatası:** Regresyon modelinin değerlendirilmesinde kullanılan en popüler metriklerden biridir. Hataların tarafsız olduğunu ve normal dağılıma sahip olduğunu varsayarak çalışır.
- **Çapraz Doğrulama:** Makine öğrenimi modelinin performansını değerlendirmek için bir başka popüler tekniktir. Bunda modeller, girdi verilerinin alt kümeleri üzerinde eğitilir ve verilerin tamamlayıcı alt kümesi üzerinde değerlendirilir.

Parametrik ve parametrik olmayan modeller

Makine öğrenmesinde temel olarak Parametrik ve Parametrik Olmayan olmak üzere iki tür model vardır. Burada parametreler, makine öğrenimi modelini oluşturmak için kullanılan tahmin değişkenleridir. Bu modellerin açıklaması aşağıda verilmiştir:

Parametrik Model: ML modelini oluşturmak için sabit sayıda parametre kullanır. **Veriler hakkında güçlü varsayımları dikkate alır.** Parametrik modellerin örnekleri, Lineer regresyon, Lojistik Regresyon, Naive Bayes, Perceptron, vb.'dir.

Parametrik Olmayan Model: Esnek sayıda parametre kullanır. **Veriler hakkında birkaç varsayımı dikkate alır. Bu modeller daha yüksek veriler için iyidir ve ön bilgi yoktur.** Parametrik olmayan modellere örnek olarak Karar Ağacı, K-En Yakın Komşu, Gauss çekirdekli SVM vb. verilebilir.

Parametrik ve parametrik olmayan modeller arasındaki farklar

Parametric model	Non-parametric model
It uses a fixed number of parameters to build the model.	It uses flexible number of parameters to build the model.
Considers strong assumptions about the data.	Considers fewer assumptions about the data.
Computationally faster	Computationally slower
Require lesser data	Require more data
Example – Logistic Regression & Naïve Bayes models	Example – KNN & Decision Tree models

Makine Öğrenmesinde Aşırı Uyum

Makine öğrenimi algoritması tüm veri noktalarını yakalamaya çalıştığında ve sonuç olarak gürültüyü de yakaladığında, modelde fazla uyum oluşur. **Bu fazla uydurma sorunu nedeniyle, algoritma düşük sapmayı, ancak çıktıda yüksek varyansı gösterir.** Aşırı uyum, makine öğrenimindeki ana sorunlardan biridir.

Veri aşırı uyum nasıl oluşur ve nasıl düzeltilebilir?

Aşırı uyum, istatistiksel bir model veya makine öğrenimi algoritması verilerin gürültüsünü yakaladığında meydana gelir. Bu, bir algoritmanın düşük yanlılık göstermesine ancak sonuçta yüksek varyans göstermesine neden olur.

Aşırı uyum, aşağıdaki metodolojiler kullanılarak önlenabilir:

ML'de Aşırı Uyumu Önleme Yöntemleri:

- Çapraz Doğrulama
- Daha fazla veri ile eğitim
- Düzenleme
- Topluluk
- Gereksiz Özelliklerin Kaldırılması
- Eğitimi Erken Durdurma.

Çapraz doğrulama: Çapraz doğrulamanın ardındaki fikir, birden çok mini tren testi bölmeleri oluşturmak için eğitim verilerini bölmektir. Bu bölmeler daha sonra modelinizi ayarlamak için kullanılabilir.

Daha fazla eğitim verisi: Makine öğrenimi modeline daha fazla veri beslemek, daha iyi analiz ve sınıflandırma yapılmasına yardımcı olabilir. Ancak, bu her zaman işe yaramaz.

Özelliklerin kaldırılması: Çoğu zaman, veri seti, analiz için gerekli olmayan alakasız özellikler veya tahmin değişkenleri içerir. Bu tür özellikler, yalnızca modelin karmaşıklığını artırır, böylece veri fazlalığı olasılıklarına yol açar. Bu nedenle, bu tür gereksiz değişkenler kaldırılmalıdır.

Erken durdurma: Bir makine öğrenimi modeli yinelemeli olarak eğitilir, bu, modelin her yinelemesinin ne kadar iyi performans gösterdiğini kontrol etmemizi sağlar. Ancak belirli sayıda yinelemeden sonra modelin performansı doygun hale gelmeye başlar. **Daha fazla eğitim, fazla uyumlu sonuçlanacaktır, bu nedenle eğitimin nerede durdurulacağını bilmek gerekir. Bu, erken durdurma adı verilen bir mekanizma ile sağlanabilir.**

Düzenleştirme (Regularization): Düzenleme n sayıda yolla yapılabilir, yöntem uyguladığınız öğrenmenin türüne bağlı olacaktır. Örneğin, karar ağaçlarında budama yapılır, bırakma tekniği sinir ağlarında kullanılır ve fazla uydurma sorunlarını çözmek için parametre ayarlama da uygulanabilir.

Topluluk modellerini kullanılması (Use Ensemble models): Topluluk öğrenimi, daha doğru sonuçlar üretmek için birleştirilen birden çok Makine Öğrenimi modeli oluşturmak için kullanılan bir tekniktir. Bu, aşırı uyumu önlemenin en iyi yollarından biridir. Bir örnek Rastgele Orman'dır, daha doğru tahminler yapmak ve fazla uydurmayı önlemek için bir karar ağaçları topluluğu kullanır.

Makine Öğrenmesinde kullanılan adımlar

Yapay zeka, denetimli ve denetimsiz öğrenme algoritmaları gibi farklı makine öğrenimi algoritmaları kullanarak dolandırıcılık tespitinde geniş ölçüde yardımcı olabilir. Makine öğreniminin kural tabanlı algoritmaları, herhangi bir işlem için kalıpları analiz etmeye ve hileli işlemleri engellemeye yardımcı olur.

Makine öğrenmesinde kullanılan adımları yazınız.

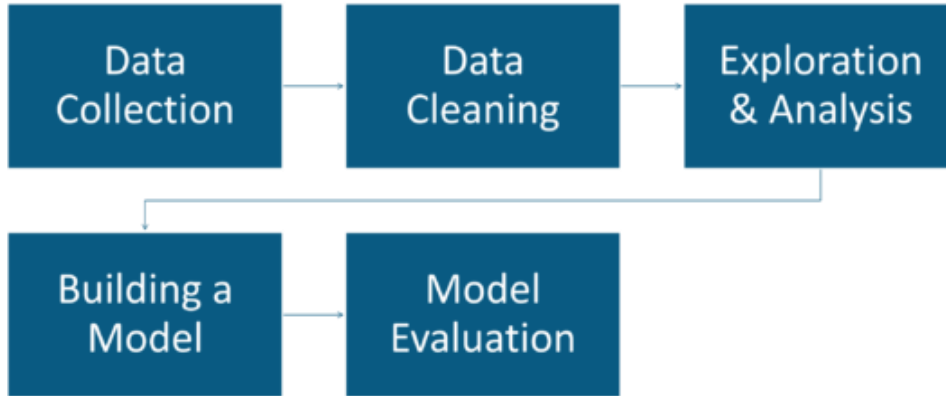
Veri Çıkarma, Veri Temizleme, Veri Keşfi ve Analizi, Öğrenen modeli oluşturma, Model Değerlendirmesi

Makine öğreniminde kullanılan adımlar aşağıda verilmiştir:

- **Veri Çıkarma (Data Extraction):** Bu aşamada ya anket yoluyla veriler toplanır ya da web kazıma yapılır. Kredi kartı sahtekarlığını tespit etmeye çalışıyorsanız, müşteriyle ilgili bilgiler toplanır. Buna işlem, alışveriş, kişisel ayrıntılar vb. dahildir. İlk adım veri çıkarmadır. Veriler bir anket yoluyla veya web kazıma araçları yardımıyla toplanır. Veri toplama, modelin türüne bağlıdır ve oluşturmak istiyoruz. Genellikle işlem detaylarını, kişisel detayları, alışverişi vb. içerir.
- **Veri Temizleme (Data Cleaning):** Bu aşamada fazlalık veriler kaldırılmalıdır. Herhangi bir tutarsızlık veya eksik değerler yanlış tahminlere yol açabilir, bu nedenle bu tür tutarsızlıklar bu adımda ele alınmalıdır. Bu adımda alakasız veya fazlalık veriler kaldırılır. Verilerde mevcut olan tutarsızlık yanlış tahminlere yol açabilmektedir.
- **Veri Keşfi ve Analizi (Data Exploration & Analysis):** Bu, yapay zekadaki en önemli adımdır. Burada çeşitli yordayıcı değişkenler arasındaki ilişkiyi incellersiniz. Örneğin, bir kişi belirli bir günde olağandışı bir miktarda para harcamışsa, dolandırıcılık olasılığı çok yüksektir. Bu tür kalıplar bu aşamada tespit edilmeli ve anlaşılmalıdır. Bu, farklı tahmin değişkenleri arasındaki ilişkiyi bulmamız gereken en önemli adımlardan biridir.
- **Bir Makine Öğrenimi modeli oluşturma (Building a Machine Learning model):** Sahtekarlığı tespit etmek için kullanılabilecek birçok makine öğrenimi algoritması vardır. Böyle bir örnek, bir sınıflandırma algoritması olan Lojistik Regresyon'dur. Olayları hileli ve hileli olmayan olmak üzere 2 sınıfa ayırmak için kullanılabilir. Model Oluşturma: Şimdi son adım, iş gereksinimine bağlı olarak farklı makine öğrenimi algoritmaları kullanarak modeli oluşturmaktır. **Regresyon veya sınıflandırma gibi.**
- **Model Değerlendirmesi (Model Evaluation):** Burada, temel olarak makine öğrenimi modelinin verimliliğini test edersiniz. İyileştirme için herhangi bir yer varsa, parametre ayarı yapılır. Bu, modelin doğruluğunu artırır.

Sahtekarlığı tespit etmede Makine Öğrenmesi Algoritmaları

Anormallikleri tespit etmek ve verilerdeki gizli kalıpları incelemek için Makine Öğrenimi algoritmaları uygulayarak Dolandırıcılık tespit problemlerinde kullanılır.



Hedef Pazar oluřturmada makine öğrenmesi

Hedef Pazarlama, bir pazarı segmentlere ayırmayı ve onu, ihtiyaç ve istekleri ürününüzle en yakından eşleşen müşterilerden oluşan birkaç ana segmente yoğunlařtırmayı içerir. Yeni işler çekmenin, satışlarınızı artırmanın ve şirketi büyütmenin anahtarıdır. Hedef pazarlamanın güzelliđi, pazarlama çabalarınızı belirli tüketici gruplarına hedefleyerek, ürünlerinizin ve/veya hizmetlerinizin tanıtımını, fiyatlandırılmasını ve dağıtımını daha kolay ve daha uygun maliyetli hale getirmesidir.



Hedeflenen pazarlamada Makine Öğrenimi:

- Metin Analitiđi Sistemleri (*Text Analytics Systems*): Metin analitiđi uygulamaları, arama uygulamalarından, metin sınıflandırmasından, adlandırılmış varlık tanımadan kalıp arama ve deđiřtirme uygulamalarına kadar uzanır.
- Kümeleme (Clustering): Müřteri segmentasyonu, hızlı arama ve görselleřtirme gibi uygulamalarla.
- Sınıflandırma (Classification): Pazarlamada metin sınıflandırması için kullanılabilen karar ağaçları ve sinir ađı sınıflandırıcıları gibi.
- Öneri Sistemleri(*Recommender Systems*): Ve pazarlama verilerinizi analiz etmek için kullanılabilecek birliktelik kuralları
- Pazar Sepeti Analizi (*Market Basket Analysis*): Pazar sepeti analizi, sık kullanılan ürünlerin kombinasyonlarını açıklar.

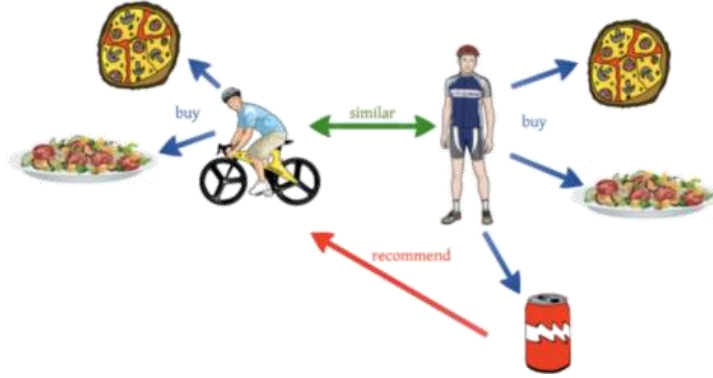
Pazar sepeti analizi

Pazar sepeti analizi, öğeler arasındaki ilişkileri bulmak için popüler bir tekniktir. Büyük perakendeciler tarafından maksimum kar elde etmek için sıklıkla kullanılır. Bu yaklaşımda, sıklıkla birlikte satın alınan ürünlerin kombinasyonlarını bulmamız gerekiyor.

Örneđin, bir kiři ekmek alırsa, tereyađı da alma řansı çok yüksektir. Bu nedenle, bu tür korelasyonları anlamak, perakendecilerin müşterilerine uygun teklifler sunarak işlerini büyötmelerine yardımcı olabilir.

'Bunu alan müşteriler bunu da aldı...' Amazon'da alışveriş yaparken bunu sıklıkla görüyoruz. Öneri motorlarının arkasındaki mantık nedir?

Amazon gibi e-ticaret siteleri, müşterilerine ürün önermek için Machine Learning'den yararlanır. Bu tür bir önerinin temel fikri, işbirlikçi filtrelemeden gelir. Ortak filtreleme, benzer alışveriş davranışına sahip yeni bir kullanıcıya ürün önermek için benzer alışveriş davranışlarına sahip kullanıcıları karşılaştırma işlemidir.



Bunu daha iyi anlamak için bir örneğe bakalım. Diyelim ki spor meraklısı bir A kullanıcısı pizza, makarna ve kola satın aldı. Şimdi birkaç hafta sonra, bisiklete binen başka bir B kullanıcısı pizza ve makarna alıyor. Kola satın almıyor, ancak alışveriş davranışları ve yaşam tarzı A kullanıcısına oldukça benzediği için Amazon B kullanıcısına bir şişe kola öneriyor. İşbirlikçi filtreleme bu şekilde çalışır.

Pazar sepeti analizinde makine öğrenmesi algoritmaları

Pazar sepeti analizi, işlemlerde sıklıkla birlikte ortaya çıkan ürün kombinasyonlarını açıklar.

Örneğin, bir kişi ekmek alırsa, tereyağı da alma olasılığı %40'tır. Ürünler arasındaki bu tür korelasyonları anlayan şirketler, bu tür ürünler üzerinde ilgili teklifler ve indirim kodları vererek işlerini büyütebilirler.

Pazar Sepeti Analizi, piyasadaki hemen hemen her büyük perakendeci tarafından takip edilen iyi bilinen bir uygulamadır. Bunun arkasındaki mantık, Association Rule Mining ve Apriori algoritması gibi Makine Öğrenimi algoritmalarıdır:

- Birliktelik kuralı madenciliği, öğelerin birbirleriyle nasıl ilişkilendirildiğini gösteren bir tekniktir.
- Apriori algoritması, birliktelik kuralları oluşturmak için sık öge kümelerini kullanır. Bir sık öge kümesinin bir alt kümesinin aynı zamanda bir sık öge kümesi olması gerektiği kavramına dayanır.



Örneğin, yukarıdaki kural, bir kişi A öğesini satın alırsa, B öğesini de satın alacağını önerir. Bu şekilde perakendeci, A ve B Öğelerini satın aldığı anda %30 indirim olacağını belirten bir indirim teklifi verebilir. C maddesine göre. Bu tür kurallar Machine Learning kullanılarak oluşturulur. Bunlar daha sonra satışları artırmak ve bir işi büyütmek için ürünlere uygulanır.

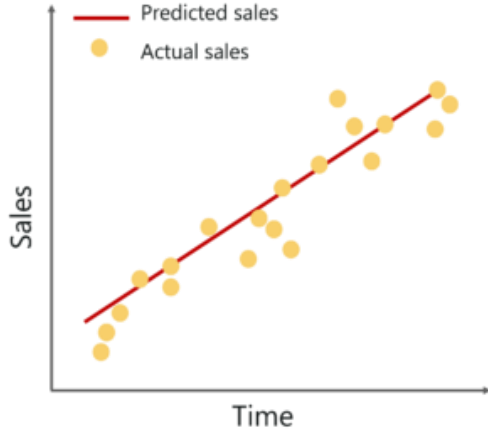
Online alışveriş işinde önümüzdeki aylara ait satışları tahmin etmek için Doğrusal Regresyon

Geçmiş verileri inceleyerek ve satışların belirli bir süre içinde nasıl değiştiğini gösteren bir model oluşturarak yapılabilir. Doğrusal Regresyon, satışları tahmin etmek için kullanılan en iyi Makine Öğrenimi algoritmalarından biridir. Hem satış hem de zaman doğrusal bir ilişkiye sahip olduğunda, basit bir doğrusal regresyon modeli kullanmak en iyisidir.

Doğrusal Regresyon, bağımsız değişkenlerin (X) değerlerine dayalı olarak bağımlı değişkeni (Y) tahmin etme yöntemidir. Bazı sürekli miktarları tahmin etmek istediğimiz durumlar için kullanılabilir.

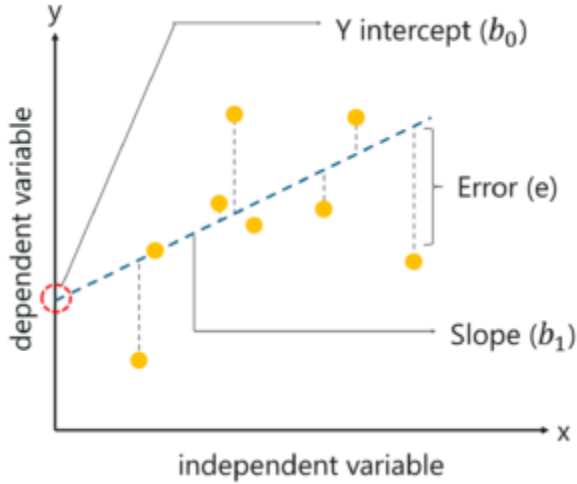
- Bağımlı değişken (Y): Değeri tahmin edilmesi gereken yanıt değişkeni.
- Bağımsız değişken (X): Yanıt değişkenini tahmin etmek için kullanılan tahmin değişkeni.

Bu örnekte, bağımlı değişken 'Y' satışları, bağımsız değişken 'X' ise süreyi temsil etmektedir. Satışlar belirli bir süre boyunca değiştiğinden, satışlar bağımlı değişkendir.



Doğrusal bir regresyon modelini temsil etmek için aşağıdaki denklem kullanılır:

$$Y = b_0 + b_1 x + e$$



Burada,

- Y = Bağımlı değişken
- b_0 = Y-Kesme
- b_1 = Çizginin eğimi
- x = Bağımsız değişken
- e = Hata

Bu nedenle, Y ekseninin satışları, X ekseninin ise dönemi ifade ettiği Lineer Regresyon modelini kullanarak, önümüzdeki aylardaki satışları kolayca tahmin edebiliriz.

Kredi onaylamada kullanılan makine öğrenmesi algoritması, KNN

Bu sorun ifadesi, başvuranın kredi talebini iki sınıfa ayıracak olan KNN algoritması kullanılarak çözülebilir:

1. Onaylandı
2. Onaylanmadı

K En Yakın Komşu, komşu veri noktalarının özelliklerine bağlı olarak yeni bir veri noktasını hedef sınıfa sınıflandıran bir Denetimli Öğrenme algoritmasıdır.



Bir kredinin onaylanması gerekip gerekmediğini tahmin etmek için aşağıdaki adımlar gerçekleştirilebilir:

Veri Çıkarma (Data Extraction): Bu aşamada ya anket yoluyla veriler toplanır ya da web kazıma yapılır. Müşteriler hakkında veriler toplanmalıdır. Buna hesap bakiyesi, kredi tutarı, yaş, meslek, kredi kayıtları vb. dahildir. Bu verileri kullanarak bir başvuru sahibinin kredisini onaylayıp onaylamayacağını tahmin edebiliriz.

Veri Temizleme (Data Cleaning): Bu aşamada gereksiz değişkenler kaldırılmalıdır. Bu değişkenlerden bazıları, örneğin Telefon, Eşzamanlı krediler vb. gibi değişkenler, bir başvuru sahibinin kredisini tahmin etmede önemli değildir. Bu tür değişkenler, yalnızca Makine Öğrenimi modelinin karmaşıklığını artıracaklarından kaldırılmalıdır.

Veri Keşfi ve Analizi (Data Exploration & Analysis): Bu, yapay zekadaki en önemli adımdır. Burada çeşitli yordayıcı değişkenler arasındaki ilişkiyi incellersiniz. Örneğin, bir kişinin ödenmemiş kredi geçmişi varsa, kredi başvurusunda bulunan kişiden onay alamama olasılığı vardır. Bu tür kalıplar bu aşamada tespit edilmeli ve anlaşılmalıdır.

Bir Makine Öğrenimi modeli oluşturma (Building a Machine Learning model): : Başvuru sahibi kredi talebinin onaylanıp onaylanmadığını tahmin etmek için kullanılacak n sayıda

makine öğrenimi algoritması vardır. Böyle bir örnek, bir sınıflandırma ve bir regresyon algoritması olan K-En Yakın Komşudur. Başvuru sahibinin kredi talebini Onaylandı ve Onaylanmadı olmak üzere iki sınıfa ayıracaktır.

Model Değerlendirmesi (Model Evaluation): Burada, temel olarak makine öğrenimi modelinin verimliliğini test edersiniz. İyileştirme için herhangi bir yer varsa, parametre ayarı yapılır. Bu, modelin doğruluğunu artırır.

Görüntü sınıflandırmasında Denetimli veya denetimsiz sınıflandırma mı? Savunun.

- Denetimli sınıflandırmada, görüntüler, özellik sınıfları oluşturmak için Makine Öğrenimi uzmanı tarafından manuel olarak beslenir ve yorumlanır.
- Denetimsiz sınıflandırmada, Makine Öğrenimi yazılımı görüntü piksel değerlerine dayalı özellik sınıfları oluşturur.

Bu nedenle, doğruluk açısından görüntü sınıflandırması için denetimli sınıflandırmayı seçmek daha iyidir.

Bilgisayarla Görme, görüntülerden veya çok boyutlu verilerden bilgi elde etmek için kullanılan bir Yapay Zeka alanıdır. Görüntü Segmentasyonu için K-means gibi Makine Öğrenimi algoritmaları kullanılır, Görüntü Sınıflandırma için Destek Vektör Makinesi gibi algoritmalar da kullanılır.



Bu nedenle Computer Vision, Nesne Algılama, Görüntü İşleme vb. gibi karmaşık sorunları çözmek için AI teknolojilerini kullanır.

Görüntü işlemedeki sonlu fark filtreleri gürültüye karşı çok hassastır. Bununla başa çıkmak için, gürültüden kaynaklanan en az bozulma olması için kullanılan yöntem

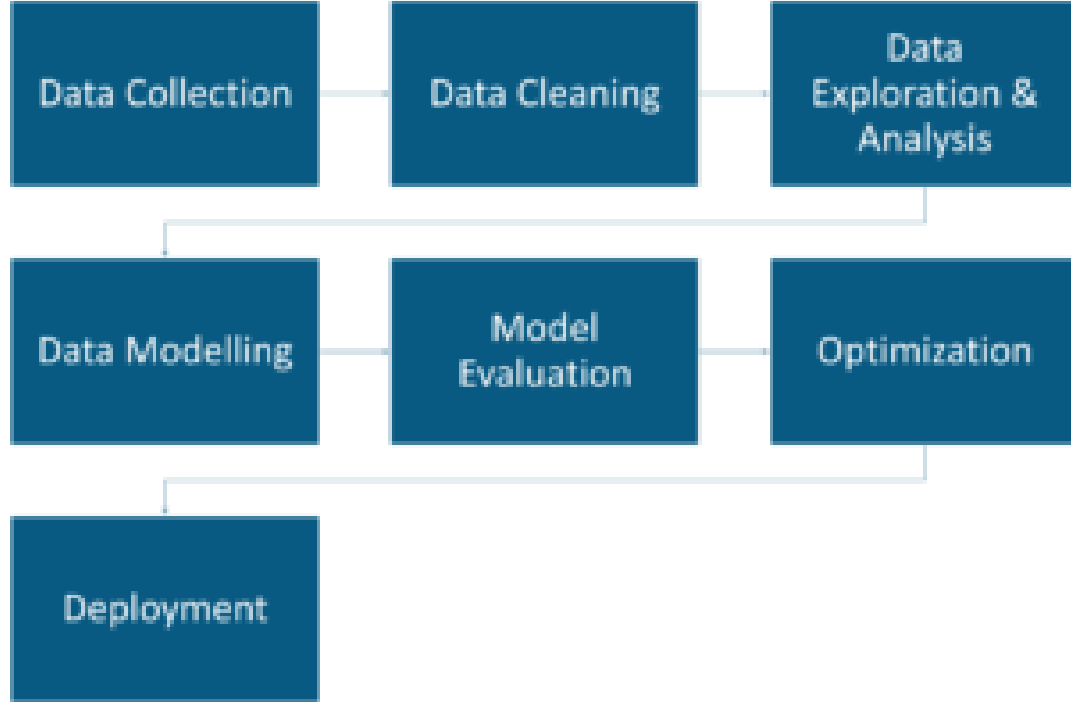
Görüntü Yumuşatma (Image Smoothing), pikselleri komşularına daha çok benzemeye zorlayarak paraziti azaltmak için kullanılan en iyi yöntemlerden biridir; bu, kontrastlardan kaynaklanan bozulmaları azaltır.



Spam mesajlarını tespit etmek ve filtrelemek için maline öğrenmesi

Spam algılamayı anlamak için Gmail örneğini ele alalım. Gmail, bu tür istenmeyen iletileri gelen kutumuzdan filtrelemek için makine öğreniminden yararlanır. Bu spam filtreleri, e-postaları spam ve spam olmayan e-postalar olmak üzere iki sınıfa ayırmak için kullanılır.

Makine öğrenimi kullanılarak spam algılamanın nasıl yapıldığını anlayalım:



- **Bir makine öğrenimi süreci her zaman veri toplama ile başlar.** Hepimiz Google'ın sahip olduğu verilerin kağıt dosyalarda olmadığını biliyoruz. Müşterinin verilerini tutan veri merkezlerine sahiptirler. E-posta içeriği, başlık, gönderen vb. veriler saklanır.
- Bunu veri temizliği takip eder. Kesin bir makine öğrenimi modeli oluşturmak için yalnızca ilgili verilerin kullanılması için gereksiz duraklama sözcüklerinden ve noktalama işaretlerinden kurtulmak önemlidir. Bu nedenle, bu aşamada 'the', 've', 'a' gibi durak sözcükleri kaldırılır. Metin, analiz edilebilecek şekilde biçimlendirilmiştir.
- **Veri temizlemeden sonra veri keşfi ve analizi gelir.** Çoğu zaman, belirli kelimeler veya ifadeler spam e-postalarda sıklıkla kullanılır. "Piyango", "kazan", "tam geri ödeme" gibi, **e-postanın spam olma olasılığının daha yüksek olduğunu gösteren kelimelere belirlenir. Bu tür kelimeler ve ortak ilişkiler bu aşamada anlaşılmalıdır.**
- Verilerden faydalı bilgiler alındıktan sonra bir makine öğrenimi modeli oluşturulur. **E-postaları istenmeyen veya istenmeyen posta olarak sınıflandırmak için Lojistik Regresyon, Naïve Bayes vb. gibi makine öğrenimi algoritmalarını kullanabilirsiniz. Makine**

öğrenimi modeli, eğitim veri kümesi kullanılarak oluşturulur. Bu veriler, modeli eğitmek ve geçmiş kullanıcı e-posta verilerini kullanarak öğrenmesini sağlamak için kullanılır.

- Bu aşamayı model değerlendirmesi takip eder. Bu aşamada model, yeni bir dizi e-postadan başka bir şey olmayan **test veri seti kullanılarak test edilir**. Bundan sonra, makine öğrenimi modeli, e-postaları doğru bir şekilde sınıflandırabilme doğruluğuna göre derecelendirilir.
- Değerlendirme sona erdikten sonra, birkaç değişken/parametre ayarlanarak modelde daha fazla iyileştirme sağlanabilir. Bu aşama aynı zamanda parametre ayarlama olarak da bilinir. Burada, temel olarak, modeli oluşturmak için kullandığınız birkaç parametreyi değiştirerek makine öğrenimi modelinin verimliliğini artırmaya çalışıyorsunuz.
- Son aşama dağıtımdır. Burada model, e-postaları gerçek zamanlı olarak işlediği ve e-postanın spam olup olmadığını tahmin ettiği son kullanıcılara dağıtılır.

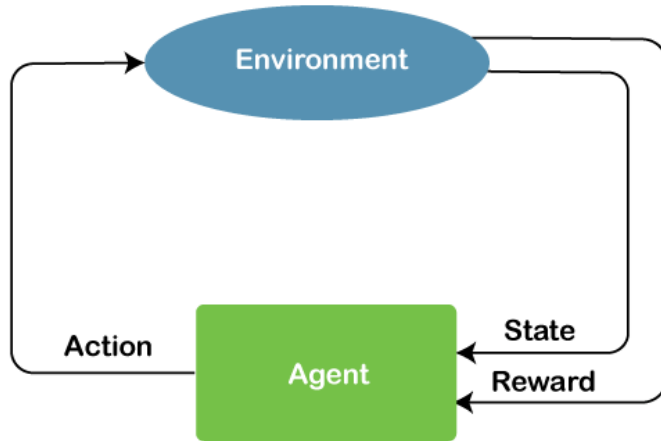
Takviyeli Öğrenme (Reinforcement Learning)

Takviyeli öğrenme, bir tür makine öğrenimidir. Bunda, bir etmen, eylemleri üreterek çevresiyle etkileşime girer ve geri bildirim yardımıyla öğrenir. Geri bildirim, temsilciye ödül şeklinde verilir, örneğin her iyi eylem için olumlu bir ödül alır ve her kötü eylem için olumsuz bir ödül alır. Acenteye herhangi bir etiketli veri veya denetim sağlanmaz. RL'de aracı, çevreyi keşfetmek için sürekli olarak üç şey yapar (eylemler gerçekleştirme, durumu değiştirme ve geri bildirim alma).

Popüler pekiştirmeli öğrenme algoritmaları şunlardır:

- Q-Öğrenme
- SARSA(Devlet Eylemi Ödül Devlet Eylemi)
- Derin Q Sinir Ağı

Takviyeli öğrenmenin çalışması aşağıdaki şema ile anlaşılabilir:



RL tabanlı sistem temel olarak aşağıdaki bileşenlerden oluşur:

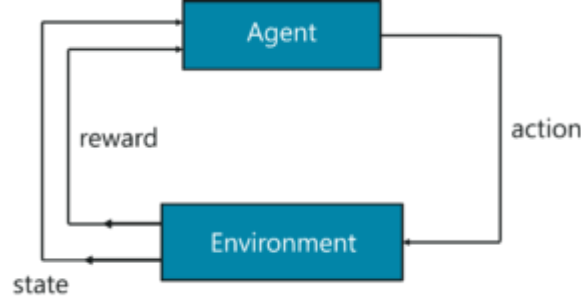
- Çevre: faktörün keşfetmesi ve üzerinde hareket etmesi gereken çevresidir.
- Faktör: Faktör, sensörlere ve aktüatörlere ve çevreyi algılama yeteneğine sahip AI programıdır.
- Durum: Çevre tarafından temsilciye iade edilen durumdur.
- Ödül: Her bir eylemi gerçekleştirdikten sonra temsilciye alınan geri bildirim.

RL'de faktör, bazı eylemler yaparak onu keşfetmek için çevre ile etkileşime girer. Her eylemde, faktörün durumu değişir veya bazen aynı kalır ve eylemin türüne göre bir ödül alır. Ödül, eyleme bağlı olarak olumsuz veya olumlu olabilen geri bildirimdir. Temsilcinin amacı, pozitif ödülü maksimize etmek ve problemin amacına ulaşmaktır.

Pekiştirmeli Öğrenme (Reinforcement Learning)

Genel olarak, bir Güçlendirme Öğrenimi (RL) sistemi iki ana bileşenden oluşur:

1. Bir temsilci
2. Bir ortam



- Ortam, aracının üzerinde işlem yaptığı ayardır ve aracı, RL algoritmasını temsil eder.
- RL süreci, ortamın aracıya bir durum gönderdiğinde başlar ve daha sonra gözlemlerine dayanarak bu duruma yanıt olarak bir eylemde bulunur.
- Sırayla, ortam bir sonraki durumu ve ilgili ödülü aracıya geri gönderir. Aracı, son eylemini değerlendirmek için bilgisini çevre tarafından döndürülen ödülle güncelleyecektir.
- Döngü, ortam uçbirim durumu gönderene kadar devam eder, bu da aracının tüm görevlerini tamamladığı anlamına gelir.

Bunu daha iyi anlamak için, temsilcimizin karşı saldırı oynamayı öğrendiğini varsayalım. RL süreci aşağıdaki adımlara ayrılabilir:

- 1) RL Aracısı (Oyuncu1) ortamdan S^0 durumunu toplar (Karşı Saldırı oyunu)
- 2) S^0 durumuna bağlı olarak, RL ajanı bir A^0 eylemi gerçekleştirir (Eylem, bir sonuca neden olan herhangi bir şey olabilir, yani ajan oyunda sola veya sağa hareket ederse). Başlangıçta, eylem rastgele
- 3) Çevre artık yeni bir S^1 durumunda (oyunda yeni aşama)
- 4) RL aracısı artık ortamdan bir R^1 ödülü alır. Bu ödül ek puan veya jeton olabilir
- 5) Bu RL döngüsü, RL aracısı ölene veya hedefe ulaşana kadar devam eder ve sürekli olarak bir durum, eylem ve ödül dizisi verir.

Q-Learning

Q-öğrenme, pekiştirmeli öğrenmede kullanılan popüler bir algoritmadır. Bellman denkleminde dayanmaktadır. Bu algortmada aracı, belirli koşullar altında ödülleri en üst düzeye çıkarmak için gerçekleştirilecek en iyi eylemleri sağlayabilecek politikaları öğrenmeye çalışır. Aracı, bu optimal politikaları geçmiş deneyimlerden öğrenir.

Q-öğrenmede, Q, her durumdaki eylemlerin kalitesini temsil etmek için kullanılır ve faktörün amacı, Q'nun değerini maksimize etmektir.

Pekiştirmeli Öğrenmede Ödül maksimizasyonun

Ödül maksimizasyonu terimi, pekiştirmeli öğrenmede kullanılır ve pekiştirmeli öğrenme faktörünün bir amacıdır. RL'de ödül, bir durumdan diğerine geçiş için harekete geçerek olumlu bir geri bildirimdir. Temsilci, optimal politikalar uygulayarak iyi bir eylem gerçekleştirirse ödül alır ve kötü bir eylem gerçekleştirirse bir ödül çıkarılır. Temsilcinin amacı, ödül maksimizasyonu olarak adlandırılan optimal politikaları uygulayarak bu ödülleri maksimize etmektir.

Explain reward maximization in Reinforcement Learning.

The RL agent works based on the theory of reward maximization. This is exactly why the RL agent must be trained in such a way that, he takes the best action so that the reward is maximum.

The collective rewards at a particular time with the respective action is written as:

$$G_t = R_{t+1} + R_{t+2} + \dots$$

Reward Maximization Equation – Artificial Intelligence Interview Questions – Edureka

The above equation is an ideal representation of rewards. Generally, things don't work out like this while summing up the cumulative rewards.



Let me explain this with a small game. In the figure you can see a fox, some meat and a tiger.

- Our RL agent is the fox and his end goal is to eat the maximum amount of meat before being eaten by the tiger.
- Since this fox is a clever fellow, he eats the meat that is closer to him, rather than the meat which is close to the tiger, because the closer he is to the tiger, the higher are his chances of getting killed.
- As a result, the rewards near the tiger, even if they are bigger meat chunks, will be discounted. This is done because of the uncertainty factor, that the tiger might kill the fox.

The next thing to understand is, how discounting of rewards work? To do this, we define a discount rate called gamma. The value of gamma is between 0 and 1. The smaller the gamma, the larger the discount and vice versa.

So, our cumulative discounted rewards is:

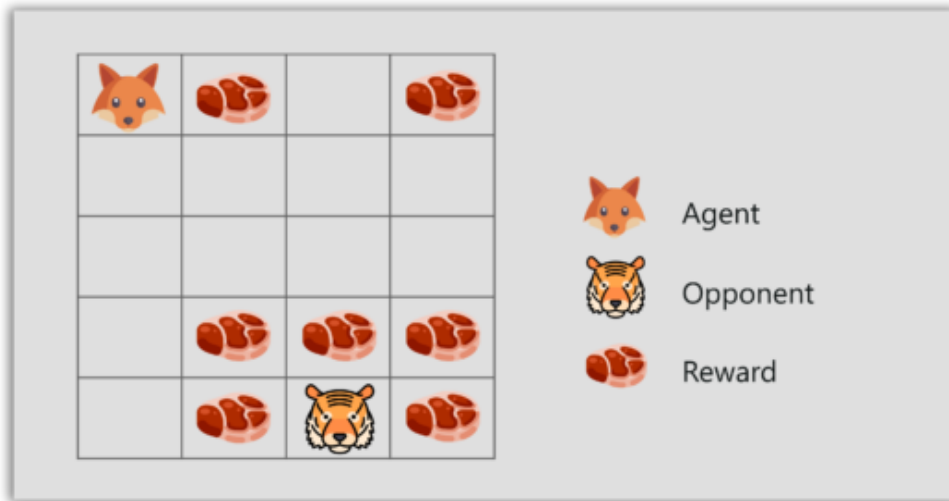
$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \text{ where } \gamma \in [0, 1)$$

$$R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \dots$$

Keşif ve keşif takası

Takviyeli öğrenmede önemli bir kavram, keşif ve kullanım değiş tokuşudur.

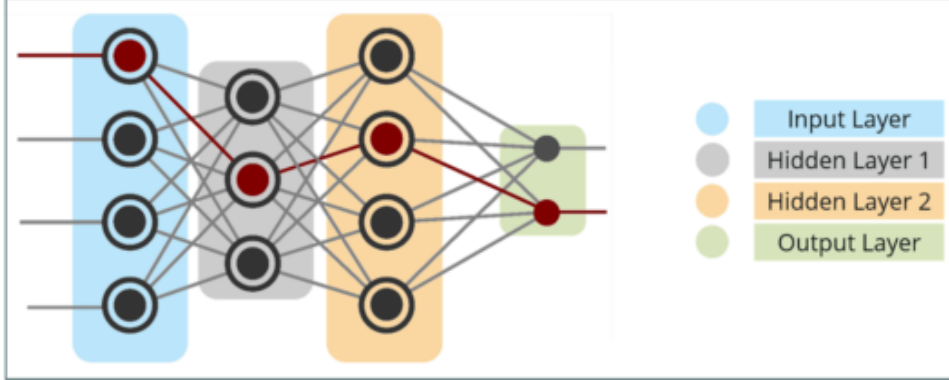
Keşif, adından da anlaşılacağı gibi, bir çevre hakkında daha fazla bilgi keşfetmek ve yakalamakla ilgilidir. Öte yandan, keşif, ödülleri artırmak için zaten bilinen keşif bilgileri kullanmakla ilgilidir.



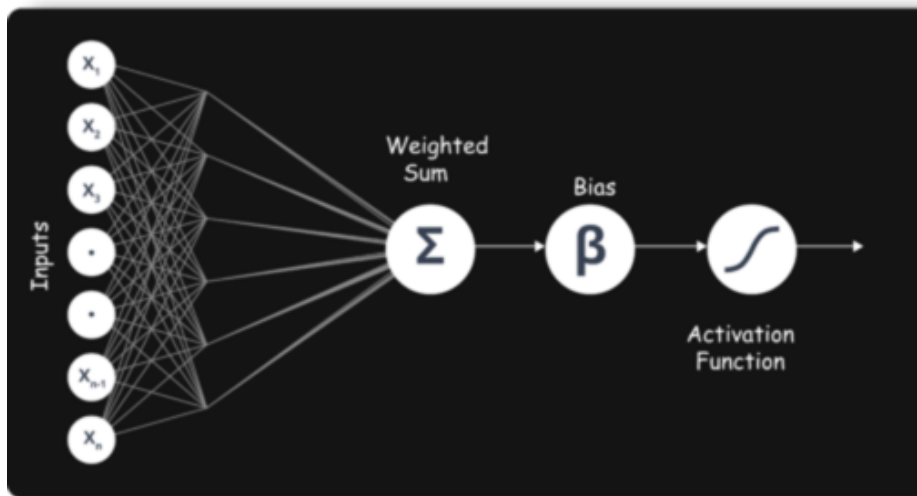
- Tilki ve kaplan örneğini düşünün, tilki sadece kendisine yakın olan et (küçük) parçalarını yer, ancak daha büyük et parçaları ona daha fazla ödül kazandırsa bile, tepedeki daha büyük et parçalarını yemez.
- Tilki sadece en yakın ödüle odaklanırsa, büyük et parçalarına asla ulaşamaz, buna keşif denir.
- Ancak tilki biraz keşfetmeye karar verirse, daha büyük ödülü, yani büyük et parçasını bulabilir. Bu keşif takasıdır.

Derin Öğrenme

Derin öğrenme, beynimizin çalışma şeklini taklit eder, yani deneyimlerden öğrenir. Karmaşık problemleri çözmek için sinir ağları kavramlarını kullanır.



- Girdi Katmanı: Bu katman tüm girdileri alır ve analiz için gizli katmana iletir.
- Gizli Katman: Bu katmanda çeşitli hesaplamalar yapılır ve sonuç çıktı katmanına aktarılır. Çözmeye çalıştığınız soruna bağlı olarak n sayıda gizli katman olabilir.
- Çıktı Katmanı: Bu katman, sinir ağından dış dünyaya bilgi aktarımından sorumludur.
- Derin Öğrenme, beyin hücresi veya nöron adı verilen beynin temel birimine dayanır. Bir nörondan esinlenerek yapay bir nöron veya bir algılayıcı geliştirildi.
- Bir biyolojik nöronun girdileri almak için kullanılan dendritleri vardır.
- Benzer şekilde, bir algılayıcı birden fazla girdi alır, çeşitli dönüşümler ve işlevler uygular ve bir çıktı sağlar.
- Tıpkı beynimizin nöral ağ adı verilen birden çok bağlantılı nöron içerdiği gibi, derin bir sinir ağı oluşturmak için algılayıcı adı verilen yapay nöronlardan oluşan bir ağa da sahip olabiliriz.



- Bir Yapay Nöron veya bir Algılayıcı, her birine belirli bir ağırlık atanmış bir dizi girdiye sahip bir nöronu modeller. Nöron daha sonra bu ağırlıklı girdiler üzerinde bazı fonksiyonları hesaplar ve çıktıyı verir.

Keras, TensorFlow ve PyTorch gibi Derin Öğrenme çerçevelerinin amacı

- Keras, Python ile yazılmış açık kaynaklı bir sinir ağı kütüphanesidir. Derin sinir ağlarıyla hızlı deneyler yapmak için tasarlanmıştır.
- TensorFlow, veri akışı programlaması için açık kaynaklı bir yazılım kitaplığıdır. Sinir ağları gibi makine öğrenimi uygulamaları için kullanılır.
- PyTorch, Python için Torch'a dayalı açık kaynaklı bir makine öğrenimi kitaplığıdır. Doğal dil işleme gibi uygulamalar için kullanılır.

Yapay Sinir Ağları

İleri Beslemeli Sinir Ağı

- Verinin veya girdinin bir yönde hareket ettiği en basit YSA biçimi.
- Veriler giriş düğümlerinden geçer ve çıkış düğümlerinden çıkar. Bu sinir ağı, gizli katmanlara sahip olabilir veya olmayabilir.

Evrişimsel Sinir Ağı

- Burada girdi özellikleri bir filtre gibi toplu olarak alınır. Bu, ağın görüntüleri parçalar halinde hatırlamasına ve işlemleri hesaplamasına yardımcı olacaktır.
- Esas olarak sinyal ve görüntü işleme için kullanılır

Tekrarlayan Sinir Ağı(RNN) – Uzun Kısa Süreli Bellek

- Bir katmanın çıktısını kaydetme ve katmanın sonucunu tahmin etmeye yardımcı olmak için bunu girdiye geri besleme ilkesine göre çalışır.
- Burada, sinir ağının ön yayılım üzerinde çalışmasına ve daha sonra kullanmak için hangi bilgilere ihtiyaç duyduğunu hatırlamasına izin verirsiniz.
- Bu şekilde her nöron bir önceki zaman adımında sahip olduğu bazı bilgileri hatırlayacaktır.

Otomatik kodlayıcılar

- Bunlar, bir girdi katmanı, bir çıktı katmanı ve bunları birbirine bağlayan bir veya daha fazla gizli katman içeren denetimsiz öğrenme modelleridir.
- Çıkış katmanı, giriş katmanı ile aynı sayıda birime sahiptir. Amacı, kendi girdilerini yeniden oluşturmaktır.
- Tipik olarak, boyut azaltma amacıyla ve üretken veri modellerini öğrenmek için.

Yaygın olarak kullanılan bazı Yapay Sinir ağları

Yapay sinir ağları, nöron adı verilen insan beyin hücrelerinin işleyişinden esinlenen istatistiksel modeldir. Bu sinir ağları, derin öğrenme ve makine öğrenimi gibi çeşitli yapay zeka teknolojilerini içerir.

Bir Yapay sinir ağı veya YSA, Girdi katmanı, Çıktı Katmanı ve gizli katmanlar dahil olmak üzere birden çok katmandan oluşur.

YSA, çeşitli derin öğrenme tekniklerinin yardımıyla örüntü tanıma, yüz tanıma vb. gibi çeşitli karmaşık sorunları çözmeye yönelik yapay zeka araçlarıdır.

Yaygın olarak kullanılan bazı Yapay sinir ağları:

- İleri Beslemeli Sinir Ağı
- Evrişimli Sinir Ağı
- Tekrarlayan Sinir Ağı
- Otomatik kodlayıcılar

Sinir ağlarında fazla takılmayı önlemek için bir teknik

Bırakma Tekniği: Bırakma tekniği, sinir ağı modellerinde fazla uydurmayı önlemek için popüler tekniklerden biridir. Eğitim sırasında rastgele seçilen nöronların bırakıldığı düzenlileştirme tekniğidir.

NLP'nin çeşitli bileşenleri

NLP, yapay zekanın bir dalı olan Doğal Dil İşleme anlamına gelir. Makinelerin insan dilini anlamasını, yorumlamasını ve manipüle etmesini sağlar. NLP'nin Bileşenleri:

Doğal Dil işlemenin temel olarak aşağıda verilen iki bileşeni vardır:

A. Doğal Dil Anlama (NLU:Natural Language Understanding):

Aşağıdaki görevleri içerir:

- a. Girdiyi faydalı temsillerle eşleştirmek için.
- b. Dilin farklı yönlerini analiz etmek.

B. Doğal Dil Üretimi (NLG: Natural Language Generation)

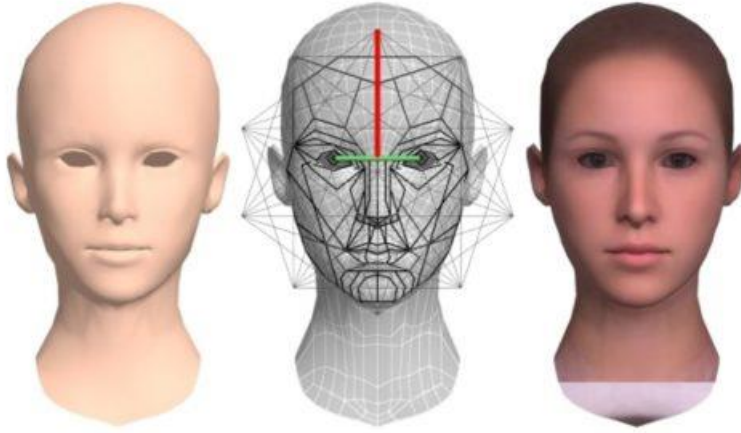
- a. Metin Planlama
- b. Cümle Planlama
- c. Metin Gerçekleştirme

Facebook tarafından kullanılan yüz tanıma algoritması

Facebook, Facebook'a bir fotoğraf yüklediğinizde size fotoğraf etiketi önerilerinin gelmesini sağlayan yüz doğrulama için derin öğrenme algoritmalarını kullanan DeepFace aracını kullanır. Derin yüz, sinir ağı modellerini kullanarak dijital görüntülerdeki yüzleri tanımlar. DeepFace'in çalışması aşağıdaki adımlarda verilmiştir:

- Önce yüklenen resimleri tarar. Görüntünün 3 boyutlu modelini yapar ve ardından bu görüntüyü farklı açılara döndürür.
- Bundan sonra eşleşmeye başlar. Bu görüntüyü eşleştirmek için, bir kişinin diğer fotoğrafları arasındaki üst düzey benzerlikleri belirlemek için bir sinir ağı modeli kullanır. Gözler arasındaki mesafe, burun şekli, göz rengi gibi farklı özellikleri kontrol eder.
- Ardından, her insan yüzü 68 belirli yüz noktasından oluştuğundan, 68 dönüm noktası testi için özyinelemeli kontrol yapar.
- Haritalama işleminden sonra görüntüyü kodlar ve o kişinin bilgilerini arar.

Facebook, yüz doğrulama için DeepFace'i kullanır. Sinir ağı modelleri kullanılarak Yapay Zeka (AI) teknikleriyle yapılandırılmış yüz doğrulama algoritması üzerinde çalışır.



Yüz doğrulama şu şekilde yapılır:

Giriş: Büyük karmaşık verilerle vahşi bir fotoğraf biçimi taranır. Bu, bulanık görüntüleri, yüksek yoğunluklu ve kontrastlı görüntüleri içerir.

Süreç: Modern yüz tanımda süreç 4 ham adımda tamamlanır:

- Yüz özelliklerini algılama
- Özellikleri hizalayın ve karşılaştırın
- 3B grafikler kullanarak anahtar kalıpları temsil edin
- Görüntüleri benzerliğe göre sınıflandırın

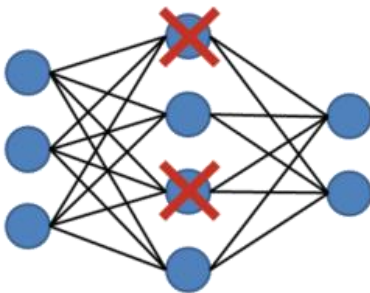
Çıktı: Nihai sonuç, 9 katmanlı derin sinir ağından türetilen bir yüz temsidir.

Eğitim Verileri: 4000'den fazla kişinin 4 milyondan fazla yüz görüntüsü

Sonuç: Facebook, iki görüntünün aynı kişiyi temsil edip etmediğini tespit edebilir

Bir sinir ağında aşırı uyumu önlemeye yardımcı olan bir teknik

Bırakma, bir sinir ağında aşırı uyumu önlemek için kullanılan bir tür düzenleme tekniğidir. Eğitim sırasında rastgele seçilen nöronların düşürüldüğü bir tekniktir.



Bir ağın Dropout (Bırakma) değeri akılcıca seçilmelidir. Çok düşük bir değer, minimum etkiye neden olur ve çok yüksek bir değer, ağı tarafından yetersiz öğrenmeye neden olur.

Derin Sinir Ağlarında hiperparametreler

Hiperparametre, tüm eğitim sürecini belirleyen ve kontrol eden parametrelerdir. Bu parametrelere örnek olarak Öğrenme hızı, Gizli Katmanlar, Gizli birimler, Etkinleştirme işlevleri vb. verilebilir. Bu parametreler modelin dışındadır. İyi hiperparametrelerin seçimi daha iyi bir algoritma yapar.

- Hiperparametreler, ağın yapısını tanımlayan değişkenlerdir. Örneğin, öğrenme oranı gibi değişkenler ağın nasıl eğitildiğini tanımlar.
- Bir ağda bulunması gereken gizli katmanların sayısını tanımlamak için kullanılırlar.
- Daha fazla gizli birim ağın doğruluğunu artırabilirken, daha az sayıda birim yetersiz donanımına neden olabilir.

Hiperparametreler ile model parametreleri arasındaki farklar

Model Parameters	Hyperparameters
Model parameters are the features of training data that will learn on its own during training.	Model Hyperparameters are the parameters that determine the entire training process.
For example, • Weights and Biases • Split points in Decision Tree	For example, • Learning Rate • Hidden Layers • Hidden Units
They are internal to the model and their value can be estimated from data.	They are external to the model and their value cannot be estimated from data.

Hiperparametre optimizasyonu için kullanılan farklı algoritmalar

Izgara Arama: Izgara arama, iki hiper parametre seti, öğrenme oranı ve katman sayısı kullanarak her kombinasyon için ağı eğitir. Daha sonra Çapraz Doğrulama tekniklerini kullanarak modeli değerlendirir.

Rastgele Arama: Arama alanını rastgele örnekler ve belirli bir olasılık dağılımından kümeleri değerlendirir. Örneğin, 10.000 örneğin tümünü kontrol etmek yerine, rastgele seçilen 100 parametre kontrol edilebilir.

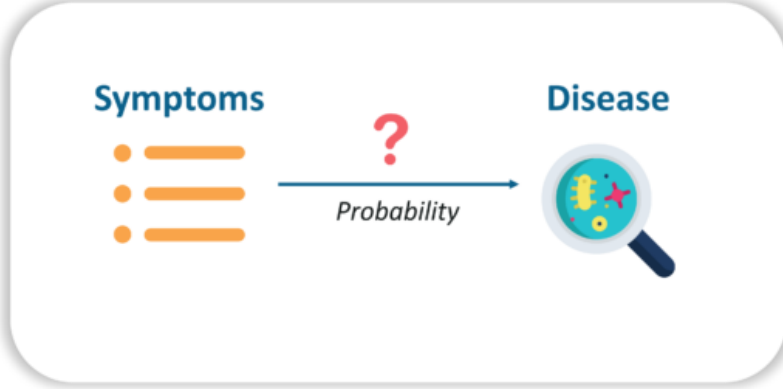
Bayes Optimizasyonu: Bu, otomatik model ayarlamayı etkinleştirerek hiperparametrelerin ince ayarını yapmayı içerir. Amaç fonksiyonuna yaklaşmak için kullanılan modele vekil model (Gauss Süreci) denir. Bayes Optimizasyonu, önceki işlemlere dayalı tahminler yapmak için sonsal işlevler elde etmek için Gauss İşlemi (GP) işlevini kullanır.

Bayes

Ağları

Bayes ağı, bir dizi değişkeni ve bunların koşullu bağımlılıklarını yönlendirilmiş bir döngüsel olmayan grafik biçiminde temsil eden istatistiksel bir modeldir.

Bir olayın meydana gelmesi üzerine, Bayes Ağları, bilinen birkaç olası nedenden herhangi birinin katkıda bulunan faktör olma olasılığını tahmin etmek için kullanılabilir.



Bayesian Network

Örneğin, hastalıklar ve semptomlar arasındaki ilişkiyi incelemek için bir Bayes ağı kullanılabilir. Çeşitli semptomlar göz önüne alındığında, Bayes ağı, çeşitli hastalıkların varlığının olasılıklarını hesaplamak için idealdir.

Bayes ağı yapay zekada neden önemlidir?

Bayes ağları, bir dizi değişken arasındaki olasılık ilişkisini göstermek için kullanılan grafik modellerdir. Birden çok kenar içeren yönlendirilmiş bir döngü grafiğidir ve her kenar koşullu bir bağımlılığı temsil eder. Bayes ağları olasılıksaldır, çünkü bu ağlar bir olasılık dağılımından oluşturulur ve ayrıca tahmin ve anomali tespiti için olasılık teorisini kullanır. Bayes teoremine dayandığından ve olasılık sorularını yanıtlamak için kullanılabildiğinden AI'da önemlidir.

Markov'un Karar süreci

Takviyeli öğrenme probleminin çözümü, Markov karar süreci veya MDP (Markov decision process) kullanılarak elde edilebilir. Bu nedenle, MDP, RL (Reinforcement Learning) problemini resmileştirmek için kullanılır. Takviyeli öğrenme problemini çözmek için matematiksel yaklaşım olarak söylenebilir. Bu sürecin temel amacı, optimum politikayı seçerek maksimum olumlu ödüller elde etmektir.

MDP'nin dört unsuru vardır:

- Sonlu durum S 'in bir A kümesi
- Sonlu eylem A 'ların bir A kümesi
- Ödüller
- Politika P_a

Bu süreçte etmen, S_1 durumundan S_2 durumuna veya başlangıç durumundan bitiş durumuna geçiş yapmak için bir A eylemi gerçekleştirir ve bu eylemleri yaparken aracı bazı ödüller alır. Aracı tarafından gerçekleştirilen eylemler dizisi politika olarak tanımlanabilir.

Özdeğerler ve özvektörler

Özvektörler ve özdeğerler, Lineer cebirin iki ana kavramıdır. Özvektörler, büyüklüğü 1.0'a eşit olan birim vektörlerdir. Özdeğerler, özvektörlere uygulanan katsayılarıdır veya bunlar özvektörün ölçeklendiği büyüklüktür.

Gizli Markov modeli

Gizli Markov modeli, bir gözlem zinciri üzerindeki olasılık dağılımlarını temsil etmek için kullanılan istatistiksel bir modeldir. Gizli markov modelinde, gizli, belirli bir zamanda üretilen bir sürecin durumunun gözlemciden gizlendiğini varsaydığı bir özelliği tanımlar ve Markov, sürecin Markov özelliğini karşıladığını varsaydığını tanımlar. HMM modelleri çoğunlukla zamansal veriler için kullanılır.

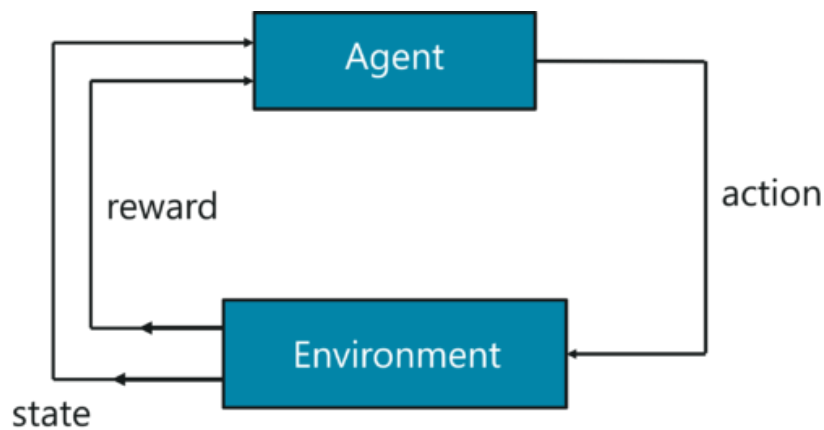
HMM, pekiştirmeli öğrenme, zamansal örüntü tanıma vb. gibi çeşitli uygulamalarda kullanılır.

Markov'un karar sürecinin bir örnekle açıklanması

The mathematical approach for mapping a solution in Reinforcement Learning is called Markov's Decision Process (MDP).

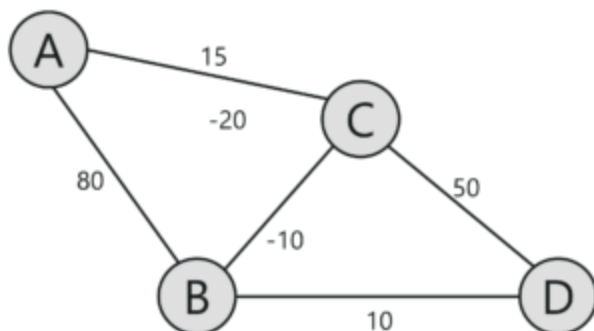
The following parameters are used to attain a solution using MDP:

- Set of actions, A
- Set of states, S
- Reward, R
- Policy, π
- Value, V



To briefly sum it up, the agent must take an action (A) to transition from the start state to the end state (S). While doing so, the agent receives rewards (R) for each action he takes. The series of actions taken by the agent, define the policy (π) and the rewards collected define the value (V). The main goal here is to maximize rewards by choosing the optimum policy.

To better understand the MDP, let's solve the Shortest Path Problem using the MDP approach:



Given the above representation, our goal here is to find the shortest path between 'A' and 'D'. Each edge has a number linked with it, this denotes the cost to traverse that edge. Now, the task at hand is to traverse from point 'A' to 'D', with minimum possible cost.

In this problem,

- The set of states are denoted by nodes i.e. {A, B, C, D}
- The action is to traverse from one node to another {A -> B, C -> D}
- The reward is the cost represented by each edge
- The policy is the path taken to reach the destination

You start off at node A and take baby steps to your destination. Initially, only the next possible node is visible to you, thus you randomly start off and then learn as you traverse through the network. The main goal is to choose the path with the lowest cost.

Since this is a very simple problem, I will leave it for you to solve. Make sure you mention the answer in the comment section.

Doğal Dil İşleme

Doğal Dil İşleme (NLP – Natural Language Processing) ve Metin madenciliği arasında ayrımlar.

Text Mining	Natural Language Processing
Aim of text mining is to extract useful insights from structured & un-structured text.	Aim of NLP is to understand what is conveyed in speech.
Text Mining can be done using text processing languages like Perl, statistical models, etc.	NLP can be achieved using advanced machine learning models, deep neural networks, etc.
Outcome: • Frequency of words • Patterns • Correlations	Outcome: • Semantic meaning of text • Sentimental analysis • Grammatical structure

NLP'nin Bileşenleri



Doğal Dil Anlayışı şunları içerir:

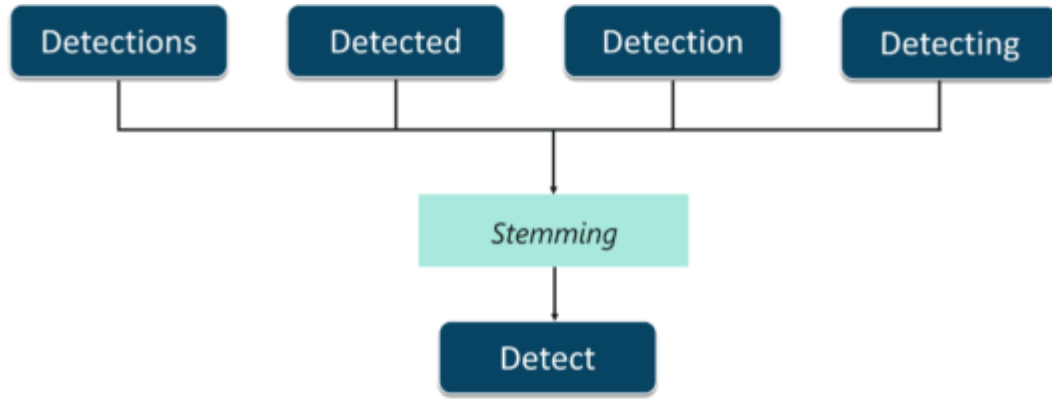
- Girdiyi faydalı temsillere eşleme
- Dilin farklı yönlerini analiz etme

Doğal Dil Üretimi şunları içerir:

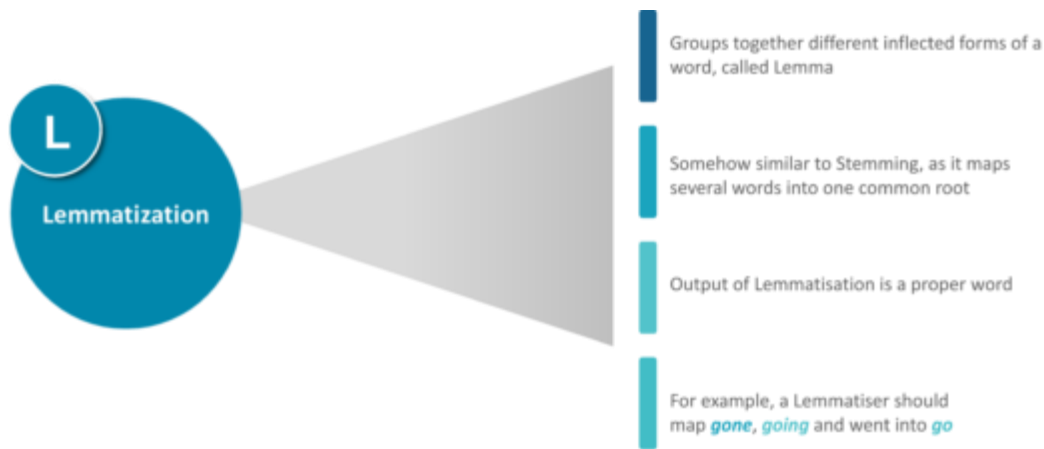
- Metin Planlama
- Cümle Planlama
- Metin Gerçekleştirme

NLP'de Stemming & Lemmatization

Stemming algoritmaları, çekimli bir kelimedeki bulunabilecek yaygın önek ve soneklerin bir listesini dikkate alarak kelimenin sonunu veya başlangıcını keserek çalışır. Bu ayırım gözetmeyen kesim bazı durumlarda başarılı olabilir, ancak her zaman değil.



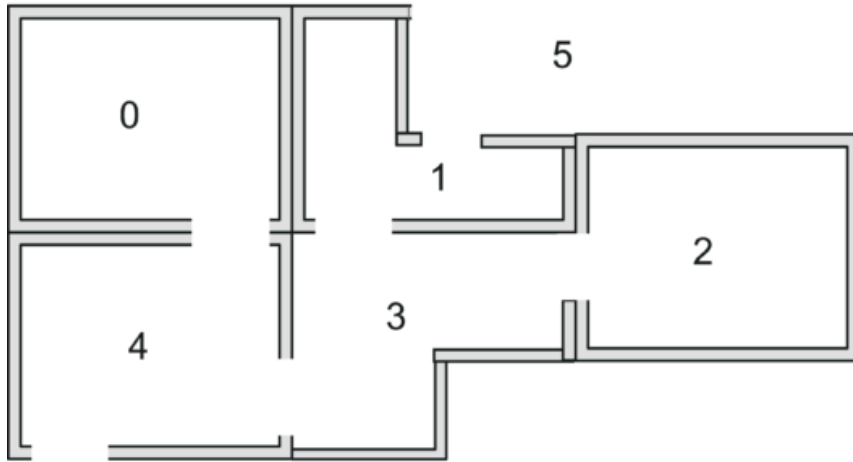
Lemmatizasyon ise kelimelerin morfolojik analizini dikkate alır. Bunu yapmak için, algoritmanın formu lemmasına geri bağlamak için inceleyebileceği ayrıntılı sözlüklere sahip olmak gerekir.



Chatbot

Chatbot, Doğal dil işlemeyi kullanarak insanlarla veya kullanıcılarla bir konuşmayı simüle edebilen Yapay zeka yazılımı veya aracıdır. Görüşme bir uygulama, web sitesi veya mesajlaşma uygulamaları aracılığıyla gerçekleştirilebilir. Bu sohbet robotları aynı zamanda dijital asistanlar olarak da adlandırılır ve insanlarla yazılı veya sesli olarak etkileşime girebilir. AI sohbet robotları, HDFC Eva sohbet botu, Vainubot vb. gibi müşterilerine 7 gün 24 saat sanal müşteri desteği sağlamak için çoğu işletmede yaygın olarak kullanılmaktadır.

Örnek: Odalardan herhangi birine (0,1,2,3,4) bir ajan yerleştirin ve amaç binanın dışına (5. oda) ulaşmaktır. Bu yapay zeka ile başarılabilir mi? Evet ise, nasıl yapılabileceğini açıklayın.

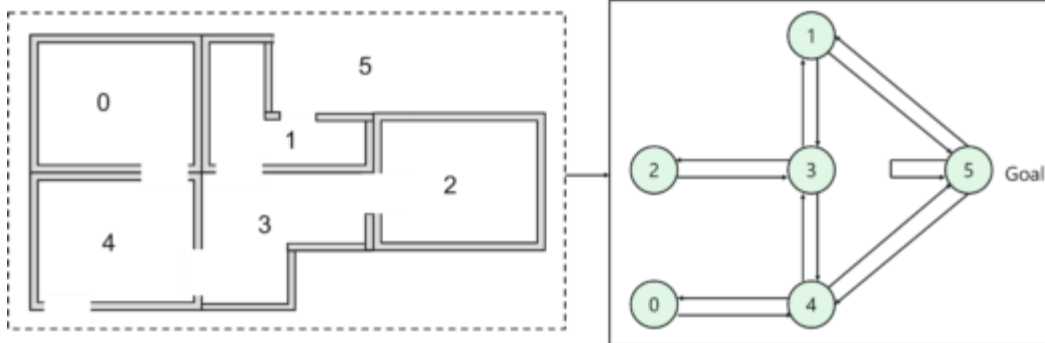


In the above figure:

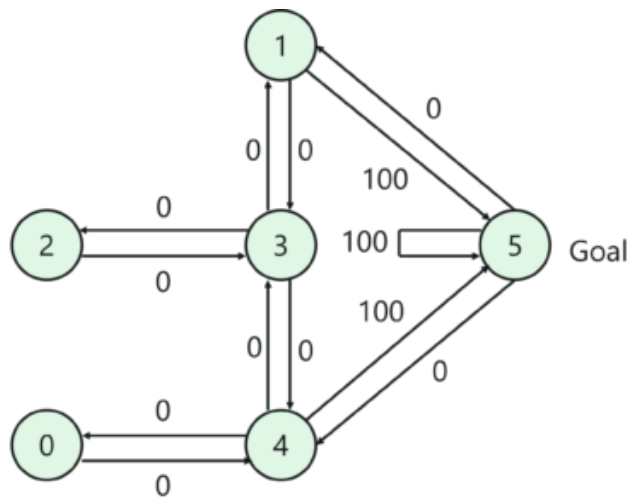
- 5 rooms in a building connected by doors
- Each room is numbered 0 through 4
- The outside of the building can be thought of as one big room (5)
- Doors 1 and 4 directly lead into the building from room 5 (outside)

This problem can be solved by using the Q-Learning algorithm, which is a reinforcement learning algorithm used to solve reward based problems.

Let's represent the rooms on a graph, each room as a node, and each door as a link, like so:



Next step is to associate a reward value to each door:



- doors that lead directly to the goal have a reward of 100
- Doors not directly connected to the target room have zero reward
- Because doors are two-way, two arrows are assigned to each room
- Each arrow contains an instant reward value

Now let's try to understand how Q-Learning can be used to solve this problem. The terminology in Q-Learning includes the terms state and action:

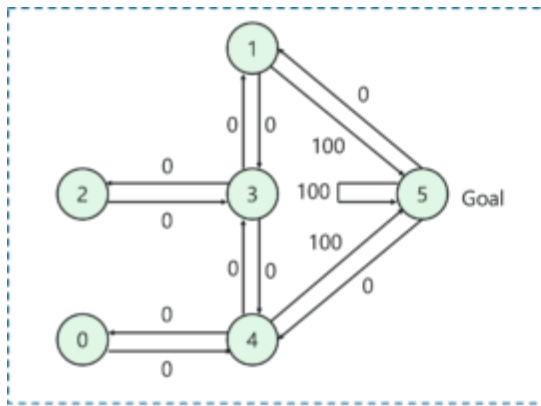
- The room (including room 5) represents a state
- Agent's movement from one room to another represents an action

In the figure, a state is depicted as a node, while "action" is represented by the arrows. Suppose, the Agent traverses from room 2 to room 5, then the following path is taken:

1. Initial state = state 2
2. State 2 -> state 3
3. State 3 -> state (2, 1, 4)
4. State 4 -> state 5

Next, we can put the state diagram and the instant reward values into a reward table or a matrix R, like so:





State	Action					
	0	1	2	3	4	5
0	-1	-1	-1	-1	0	-1
1	-1	-1	-1	0	-1	100
2	-1	-1	-1	0	-1	-1
3	-1	0	0	-1	0	-1
4	0	-1	-1	0	-1	100
5	-1	0	-1	-1	0	100

The -1's in the table represent null values

The next step is to add another matrix Q, representing the memory of what the agent has learned through experience.

- The rows of matrix Q represent the current state of the agent
- columns represent the possible actions leading to the next state

The formula to calculate the Q matrix:

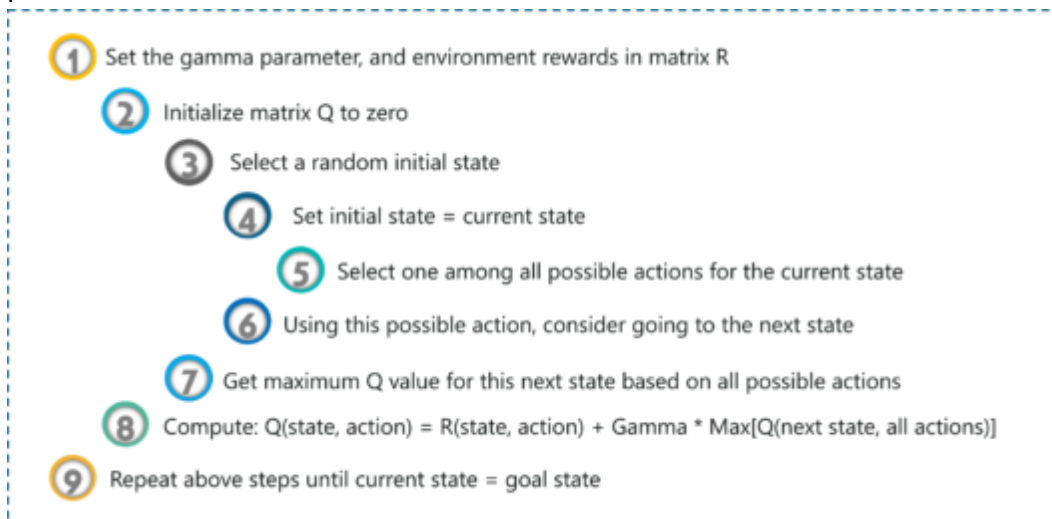
$$Q(\text{state}, \text{action}) = R(\text{state}, \text{action}) + \text{Gamma} * \text{Max}[Q(\text{next state}, \text{all actions})]$$

Here, Q(state, action) and R(state, action) represent the state and action in the Reward matrix R and the Memory matrix Q.

Note: The Gamma parameter has a range of 0 to 1 ($0 \leq \text{Gamma} < 1$).

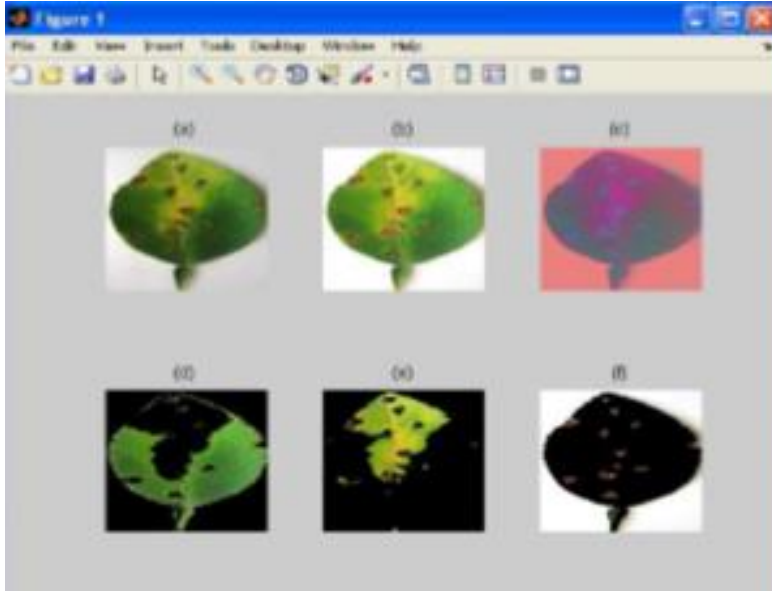
- If Gamma is closer to zero, the agent will tend to consider only immediate rewards.
- If Gamma is closer to one, the agent will consider future rewards with greater weight

Finally, by following the below steps, the agent will reach room 5 by taking the most optimal path:



Örnek: Hindistan'da mahsul verimi düşüyor çünkü çiftçiler mahsullerdeki hastalıkları erken aşamalarda tespit edemiyorlar. Makine Öğrenmesi, mahsullerde hastalık tespiti için kullanılabilir mi? Öyleyse, açıklayın.

Makine Öğrenmesi, yaprak hastalıklarının çıkarılması ve sınıflandırılması için görüntü işleme ve sınıflandırma tekniklerini uygulamak için kullanılabilir.



Karmaşıklığı önlemek için adımlara ayırılım:

Görüntü Toplama: Örnek görüntüler toplanır ve bir girdi veri tabanı olarak saklanır.

Görüntü Ön İşleme: Görüntü ön işleme aşağıdakileri içerir:

- İstenmeyen bozulmayı önleyen görüntü verileri iyileştirilir.
- Görüntü özellikleri geliştirilir.
- Görüntü kırpma, geliştirme, renk alanı dönüştürülür.
- Bir görüntünün kontrastını ayarlamak için Histogram eşitleme gerçekleştirilir.

Görüntü Segmentasyonu: Sayısal bir görüntünün, görüntü analizinin kolaylaşması için birden çok parçaya bölünmesi işlemidir. Segmentasyon, renk, doku gibi görüntü özelliklerine dayanmaktadır. Segmentasyon için kullanılan popüler bir Makine Öğrenimi yöntemi, K-means kümeleme algoritmasıdır.

Özellik Çıkarımı: Bu, belirli bir örneğin önemini bulmak için kullanılabilecek bilgileri çıkarmak için yapılır. Haar Wavelet dönüşümü doku analizi için kullanılabilir ve hesaplamalar Gray-Level Co-Occurrence Matrix kullanılarak yapılabilir.

Sınıflandırma: Son olarak, yaprak hastalıklarının sınıflandırılması için Lineer Destek Vektör Makinesi kullanılmaktadır. SVM, iki sınıf arasındaki karar sınırı olarak adlandırılan bir hiper düzlem kullanan ikili bir sınıflandırıcıdır. Bu, iki sınıfın oluşumuyla sonuçlanır:

1. Hastalıklı yapraklar

2. Sağlıklı yapraklar

Bu nedenle, AI, görüntüleri inceleyerek ve işleyerek hastalıkları sınıflandırmak ve tespit etmek için Bilgisayarla Görmede kullanılabilir. Bu, yapay zekanın en derin uygulamalarından biridir.

Sorular

Q1. You are part of data science team that is working for a national fast-food chain. You create a simple report that shows trend: Customers who visit the store more often and buy smaller meals spend more than customers who visit less frequently and buy larger meals. What is the most likely diagram that your team created?

Ulusal bir fast food zinciri için çalışan veri bilimi ekibinin bir parçasısınız. Trendi gösteren basit bir rapor oluşturun: Mağazayı daha sık ziyaret eden ve daha küçük öğünler satın alan müşteriler, daha az sıklıkta ziyaret eden ve daha büyük öğünler satın alan müşterilerden daha fazla harcama yapar. Ekibinizin oluşturduğu en olası diyagram nedir?

- multiclass classification diagram
- **linear regression and scatter plots**
- pivot table
- K-means cluster diagram

Sınıflandırma için Naïve Bayes, Karar ağaçları, SVM, Random Forest sınıflandırıcı, KNN, lojistik regresyon gibi birçok algoritma kullanıyoruz. Dengesiz veri kümesine sahip birden fazla sınıfa sahip sınıflandırma problemleri, ikili bir sınıflandırma probleminden farklı bir zorluk sunar. Çarpık dağılım, birçok geleneksel makine öğrenimi algoritmasını, özellikle azınlık sınıfı örneklerini tahmin etmede daha az etkili hale getirir. **Çok Sınıflı Sınıflandırma:** İki'den fazla sınıfı olan bir sınıflandırma görevi; örneğin, portakal, elma veya armut olabilecek bir dizi meyve görüntüsünü sınıflandırılmasıdır. Çok sınıflı sınıflandırma, her örneğin bir ve yalnızca bir etikete atandığı varsayımını yapar: bir meyve, elma veya armut olabilir, ancak ikisi aynı anda olamaz.

Dağılım diyagramı, iki değişken arasındaki ilişkiyi göstermek için kullanılan son derece basit bir istatistiksel araçtır. Genellikle iki değişken arasına bir model uydurmak için kullanılan basit bir doğrusal regresyon çizgisi ile birleştirilir.

Pandalardaki özet tablo (**pivot table**), diğer iki kategorik değişkene dayalı olarak bir veya daha fazla sayısal değişkeni özetlemek için mükemmel bir araçtır. DataFrame olarak elektronik tablo stilinde bir pivot tablo oluşturun. Pandaların pivot tablosundaki seviyeler, DataFrame sonucunun indeksi ve sütunlarındaki Çoklu indeks nesnelerinde (hiyerarşik indeksler) saklanacaktır.

K-Means kümeleme, mevcut girdi verilerinin etiketli bir yanıt olmadığı denetimsiz algoritmalarından biridir. K-Means kümeleme, denetimsiz bir öğrenme algoritmasıdır. Denetimli öğrenmenin aksine, bu kümeleme için etiketlenmiş veri yoktur. K-Means, nesnelerin benzerlerini paylaşan ve başka bir kümeye ait nesnelere benzemeyen kümelere bölünmesini gerçekleştirir.

Q2. You work for an organization that sells a spam filtering service to large companies. Your organization wants to transition its product to use machine learning. It currently a list Of 250,00 keywords. If a message contains more than few of these keywords, then it is identified as spam. What would be one advantage of transitioning to machine learning?

- The product would look for new patterns in spam messages.
- The product could go through the keyword list much more quickly.
- The product could have a much longer keyword list.
- **The product could find spam messages using far fewer keywords.**

Büyük şirketlere spam filtreleme hizmeti satan bir kuruluştaki çalışıyorsunuz. Kuruluşunuz, ürününü makine öğrenimini kullanacak şekilde geçirmek istiyor. Şu anda 250,00 anahtar kelimeden oluşan bir liste. Bir ileti bu anahtar kelimelerden birkaçından fazlasını içeriyorsa, spam olarak tanımlanır. Makine öğrenimine geçmenin avantajlarından biri ne olabilir?

- Ürün, spam iletilerinde yeni kalıplar arar.
- Ürün, anahtar kelime listesinden çok daha hızlı geçebilir.
- Ürün çok daha uzun bir anahtar kelime listesine sahip olabilir.
- **Ürün, çok daha az anahtar kelime kullanarak spam iletileri bulabilir.**

Q3. You work for a music streaming service and want to use supervised machine learning to classify music into different genres. Your service has collected thousands of songs in each genre, and you used this as your training data. Now you pull out a small random subset of all the songs in your service. What is this subset called?

- data cluster
- Supervised set
- big data
- **test data**

Bir müzik akışı hizmeti için çalışıyorsunuz ve müziği farklı türlere göre sınıflandırmak için denetimli makine öğrenimini kullanmak istiyorsunuz. Hizmetiniz her türde binlerce şarkı topladı ve siz bunu eğitim veriniz olarak kullandınız. Şimdi hizmetinizdeki tüm şarkıların rastgele küçük bir alt kümesini çıkarırsınız. Bu alt kümeye ne denir?

Q4. In traditional computer programming, you input commands. What do you input with machine learning?

- patterns
- programs
- rules
- **data**

Geleneksel bilgisayar programcılığında komutları girersiniz. Makine öğrenimi ile ne giriyorsunuz?

Q5. Your company wants to predict whether existing automotive insurance customers are more likely to buy homeowners insurance. It created a model to better predict the best customers contact about homeowners insurance, and the model had a low variance but high bias. What does that say about the data model?

- **It was consistently wrong.**
- It was inconsistently wrong.
- It was consistently right.
- It was equally right and wrong.

Şirketiniz, mevcut otomotiv sigortası müşterilerinin ev sahibi sigortası satın alma olasılığının daha yüksek olup olmadığını tahmin etmek istiyor. Ev sahibi sigortası hakkında en iyi müşteri temasını daha iyi tahmin etmek için bir model oluşturdu ve **modelin varyansı düşük, ancak yanlılığı yüksekti**. Bu, veri modeli hakkında ne denir?

- **Sürekli yanlıştı.**
- Tutarsız bir şekilde yanlıştı.
- Sürekli olarak doğruydı.
- Aynı derecede doğru ve yanlıştı.

Q6. You want to identify global weather patterns that may have been affected by climate change. To do so, you want to use machine learning algorithms to find patterns that would otherwise be imperceptible to a human meteorologist. What is the place to start?

- Find labeled data of sunny days so that the machine will learn to identify bad weather.
- **Use unsupervised learning have the machine look for anomalies in a massive weather database.**
- Create a training set of unusual patterns and ask the machine learning algorithms to classify them.
- Create a training set of normal weather and have the machine look for similar patterns.

İklim değişikliğinden etkilenmiş olabilecek küresel hava durumu modellerini belirlemek istiyorsunuz. Bunu yapmak için, bir insan meteoroloğu tarafından algılanamayacak olan kalıpları bulmak için makine öğrenimi algoritmalarını kullanmak istiyorsunuz. Başlamak için en iyi yer neresidir?

- Güneşli günlerin etiketli verileri bulunur, böylece makine kötü havayı tanımlamayı öğrenecektir.
- **Denetimsiz öğrenme kullanılır, makinenin devasa bir hava durumu veritabanındaki anormallikleri araması sağlanır.**
- Alışılmadık kalıplardan oluşan bir eğitim seti oluşturulur ve makine öğrenimi algoritmalarından bunları sınıflandırmasını istenir.
- Normal hava koşullarından oluşan bir eğitim seti oluşturulur ve makinenin benzer kalıpları araması sağlanır.

Q7. You work in a data science team that wants to improve the accuracy of its K-nearest neighbor result by running on top of a naive Bayes result. What is this an example of?

- regression
- boosting
- bagging
- **stacking**

Q8. ____ looks at the relationship between predictors and your outcome (tahmin ediciler ve sonucunuz arasındaki ilişki).

- **Regression analysis**
- K-means clustering
- Big data
- Unsupervised learning

Regresyon analizi, iki ya da daha çok nicel değişken arasındaki ilişkiyi ölçmek için kullanılan analiz metodudur.

K-araç kümeleme, başlangıçta sinyal işlemeden gelen, n gözlemi, her gözlemin en yakın ortalamaya sahip kümeye (küme merkezleri veya küme merkezi) ait olduğu k kümeye bölmeyi amaçlayan bir vektör niceleme yöntemidir.

Gözetimsiz öğrenme, gözetimli öğrenmeden farklı olarak, verileri sebep-sonuç ya da giriş-çıkış şeklinde etiketlemeden, veri içerisinde var olan ilişkilerin ve yapıların öğrenilmesidir.

Q9. What is an example of a commercial application for a machine learning system?

a data entry system

a data warehouse system

a massive data repository (büyük bir veri deposu)

a product recommendation system

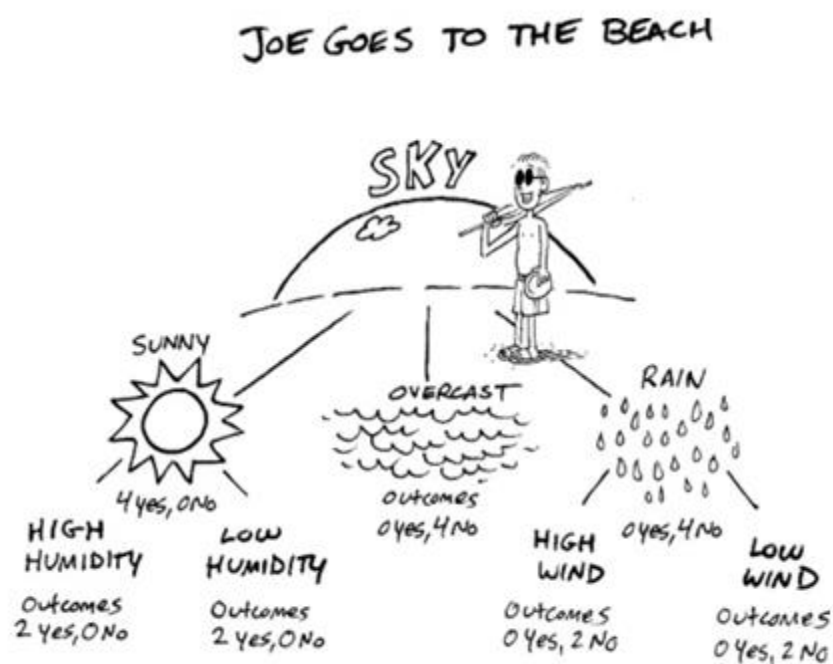
Q10. What does this image illustrate?

a decision tree

reinforcement learning

K-nearest neighbor

a clear trendline



Q11. You work for a power company that owns hundreds of thousands of electric meters. These meters are connected to the internet and transmit energy usage data in real-time. Your supervisor asks you to direct project to use machine learning to analyze this usage data. Why are machine learning algorithms ideal in this scenario?

- The algorithms would help the meters access the internet.
- The algorithms will improve the wireless connectivity.
- **The algorithms would help your organization see patterns of the data.**
- By using machine learning algorithms, you are creating an IoT device.

Yüz binlerce elektrik sayacına sahip bir elektrik şirketinde çalışıyorsunuz. Bu sayaçlar internete bağlıdır ve enerji kullanım verilerini gerçek zamanlı olarak iletir. Yöneticiniz, bu kullanım verilerini analiz etmek için makine öğrenimini kullanmak üzere projeyi yönlendirmenizi ister. Bu senaryoda makine öğrenimi algoritmaları neden idealdir?

- Algoritmalar, sayaçların internete erişmesine yardımcı olacaktır.
- Algoritmalar kablosuz bağlantıyı geliştirecektir.
- **Algoritmalar, kuruluşunuzun veri kalıplarını görmesine yardımcı olur.**
- Makine öğrenimi algoritmalarını kullanarak bir IoT cihazı oluşturuyorsunuz.

Q12. To predict a quantity value. use ____.

- **regression**
- clustering
- classification
- dimensionality reduction

Q13. Why is naive Bayes called naive?

- It naively assumes that you will have no data.
- It does not even try to create accurate predictions.
- **It naively assumes that the predictors are independent from one another.**
- It naively assumes that all the predictors depend on one another.

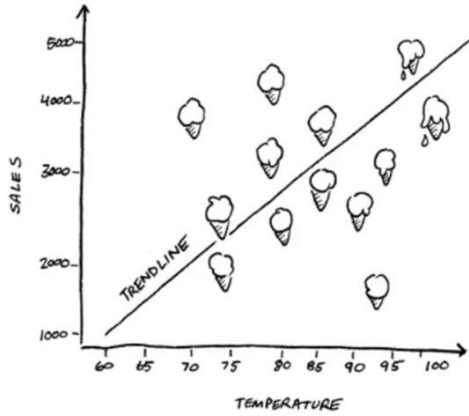
Naive Bayes'e neden saf denir?

- Saf bir şekilde hiçbir veriye sahip olmayacağını varsayar.
- Doğru tahminler oluşturmaya bile çalışmaz.
- **Tahmin edicilerin birbirinden bağımsız olduğunu saf bir şekilde varsayar.**
- Saf bir şekilde tüm tahmin edicilerin birbirine bağlı olduğunu varsayar.

Açıklama: Naive Bayes, her bir girdi değişkeninin bağımsız olduğunu varsaydığı için saf olarak adlandırılır. Bu güçlü bir varsayımdır ve gerçek veriler için gerçekçi değildir; bununla birlikte, teknik çok çeşitli karmaşık problemler üzerinde çok etkilidir.

Q14. You work for an ice cream shop and created the chart below, which shows the relationship between the outside temperature and ice cream sales. What is the best description of this chart?

- **It is a linear regression chart.**
- It is a supervised trendline chart.
- It is a decision tree.
- It is a clustering trend chart.



Bir dondurma dükkanında çalışıyorsunuz ve dış hava sıcaklığı ile dondurma satışları arasındaki ilişkiyi gösteren aşağıdaki grafiği oluşturdunuz. Bu grafiğin en iyi açıklaması nedir?

- **Doğrusal bir regresyon tablosudur.**
- Denetimli bir trend çizgisi grafiğidir.
- Bir karar ağacıdır.
- Bir kümelenme trend grafiğidir.

Q15. How is machine learning related to artificial intelligence?

- Artificial intelligence focuses on classification, while machine learning is about clustering data.
- **Machine learning is a type of artificial intelligence that relies on learning through data.**
- Artificial intelligence is form of unsupervised machine learning.
- Machine learning and artificial intelligence are the same thing.

Yapay zeka ile makine öğrenimi arasında nasıl bir ilişki var?

- Yapay zeka sınıflandırmaya odaklanırken, makine öğrenimi verileri kümeleme ile ilgilidir.
- **Makine öğrenimi, veriler yoluyla öğrenmeye dayanan bir tür yapay zekadır.**
- Yapay zeka, denetimsiz makine öğrenimi biçimidir.
- Makine öğrenimi ve yapay zeka aynı şeydir.

Q16. How do machine learning algorithms make more precise predictions?

- The algorithms are typically run more powerful servers.
- **The algorithms are better at seeing patterns in the data.**
- Machine learning servers can host larger databases.
- The algorithms can run on unstructured data.

Makine öğrenimi algoritmaları nasıl daha kesin tahminler yapar?

- Algoritmalar genellikle daha güçlü sunucularda çalıştırılır.
- **Algoritmalar, verilerdeki kalıpları görmede daha iyidir.**
- Makine öğrenimi sunucuları daha büyük veritabanlarını barındırabilir.
- Algoritmalar yapılandırılmamış veriler üzerinde çalışabilir.

Q17. You work for an insurance company. Which machine learning project would add the most value for the company!

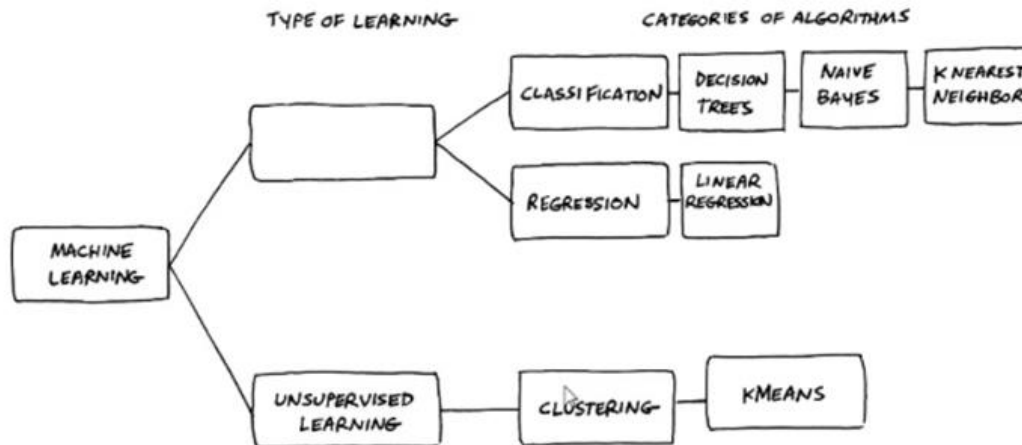
- Create an artificial neural network that would host the company directory.
- **Use machine learning to better predict risk.**
- Create an algorithm that consolidates all of your Excel spreadsheets into one data lake.
- Use machine learning and big data to research salary requirements.

Bir sigorta şirketinde çalışıyorsun. Hangi makine öğrenimi projesi şirket için en fazla değeri katacak!

- Şirket dizinini barındıracak bir yapay sinir ağı oluşturulur.
- **Riski daha iyi tahmin etmek için makine öğrenimini kullanılır.**
- Tüm Excel elektronik tablolarınızı tek bir veri gölünde birleştiren bir algoritma oluşturulur.
- Maaş gereksinimlerini araştırmak için makine öğrenimini ve büyük verileri kullanılır.

Q18. What is the missing information in this diagram?

- Training Set
- Unsupervised Data
- **Supervised Learning**
- Binary Classification



Q19. What is one reason not to use the same data for both your training set and your testing set?

- You will almost certainly underfit the model.
- You will pick the wrong algorithm.
- You might not have enough data for both.
- **You will almost certainly overfit the model.**

Hem eğitim setiniz hem de test setiniz için aynı verileri kullanmamanın bir nedeni nedir?

- Neredeyse kesinlikle modele uyacaksınız.
- Yanlış algoritmayı seçeceksin.
- Her ikisi için de yeterli veriniz olmayabilir.
- **Neredeyse kesinlikle modele fazla uyacaksınız.**

Q20. Your university wants to use machine learning algorithms to help sort through incoming student applications. An administrator asks if the admissions decisions might be biased against any particular group, such as women. What would be the best answer?

- Machine learning algorithms are based on math and statistics, and so by definition will be unbiased.
- There is no way to identify bias in the data.
- Machine learning algorithms are powerful enough to eliminate bias from the data.
- **All human-created data is biased, and data scientists need to account for that.**

Explanation: While machine learning algorithms don't have bias, the data can have them.

Üniversiteniz, gelen öğrenci uygulamalarını sıralamaya yardımcı olması için makine öğrenimi algoritmalarını kullanmak istiyor. Bir yönetici, kabul kararlarının kadınlar gibi belirli bir gruba karşı yanlı olup olmadığını sorar. En iyi cevap ne olurdu?

- Makine öğrenimi algoritmaları matematik ve istatistiğe dayalıdır ve bu nedenle tanım gereği tarafsız olacaktır.
- Verilerdeki yanlılığı tanımlamanın bir yolu yoktur.
- Makine öğrenimi algoritmaları, verilerdeki yanlılığı ortadan kaldıracak kadar güçlüdür.
- **İnsan tarafından yaratılan tüm veriler yanlıdır ve veri bilimcilerinin bunu hesaba katması gerekir.**

Açıklama: Makine öğrenimi algoritmalarında yanlılık bulunmasa da verilerde yanlılık olabilir.

Q21. What is stacking?

- **The predictions of one model become the inputs another.**
- You use different versions of machine learning algorithms.
- You use several machine learning algorithms to boost your results.
- You stack your training set and testing set together.

Yığınlama nedir?

- **Bir modelin tahminleri diğerinin girdileri haline gelir.**
- Makine öğrenimi algoritmalarının farklı sürümlerini kullanıyorsunuz.
- Sonuçlarınızı artırmak için birkaç makine öğrenimi algoritması kullanırsınız.
- Eğitim setinizi ve test setinizi bir araya toplarsınız.

Q22. You want to create a supervised machine learning system that identifies pictures of kittens on social media. To do this, you have collected more than 100,000 images of kittens. What is this collection of images called?

- **training data**
- linear regression
- big data
- test data

Sosyal medyadaki yavru kedi resimlerini tanımlayan denetimli bir makine öğrenimi sistemi oluşturmak istiyorsunuz. Bunu yapmak için 100.000'den fazla yavru kedi resmi topladınız. Bu resim koleksiyonuna ne ad verilir?

- **Eğitim verileri**
- doğrusal regresyon
- Büyük veri
- test verisi

Q23. You are working on a project that involves clustering together images of different dogs. You take image and identify it as your centroid image. What type machine learning algorithm are you using?

- centroid reinforcement
- K-nearest neighbor
- binary classification
- **K-means clustering**

Explanation: The problem explicitly states "clustering".

Farklı köpeklerin görüntülerini bir araya getirmeyi içeren bir proje üzerinde çalışıyorsunuz. Görüntüyü çeker ve onu merkez görüntünüz olarak tanımlarsınız. Ne tür makine öğrenme algoritması kullanıyorsunuz?

- ağırlık merkezi takviyesi
- K-en yakın komşu
- ikili sınıflandırma
- **K-kümeleme anlamına gelir**

Açıklama: Sorun açıkça "kümeleme"yi belirtiyor.

K-ortalama kümeleme ya da K-means kümeleme (K-means clustering) yöntemi N adet veri nesnesinden oluşan bir veri kümesini giriş parametresi olarak verilen K adet kümeye bölümlenektir. Amaç, gerçekleştirilen bölümlenme işlemi sonunda elde edilen kümelerin, küme içi benzerliklerinin maksimum ve kümeler arası benzerliklerinin ise minimum olmasını sağlamaktır.

Q24. Your company wants you to build an internal email text prediction model to speed up the time that employees spend writing emails. What should you do?

- **Include training email data from all employees.**
- Include training email data from new employees.
- Include training email data from seasoned employees.
- Include training email data from employees who write the majority of internal emails.

Şirketiniz, çalışanların e-posta yazmak için harcadıkları zamanı hızlandırmak için dahili bir e-posta metni tahmin modeli oluşturmanızı istiyor. Ne yapmalısınız?

- **Tüm çalışanlardan gelen eğitim e-posta verilerini dahil edin.**
- Yeni çalışanlardan gelen eğitim e-posta verilerini ekleyin.
- Deneyimli çalışanlardan gelen eğitim e-posta verilerini ekleyin.
- Dahili e-postaların çoğunu yazan çalışanlardan gelen eğitim e-posta verilerini ekleyin.

Q25. Your organization allows people to create online professional profiles. A key feature is the ability to create clusters of people who are professionally connected to one another. What type of machine learning method is used to create these clusters?

- **unsupervised machine learning**
- binary classification
- supervised machine learning
- reinforcement learning

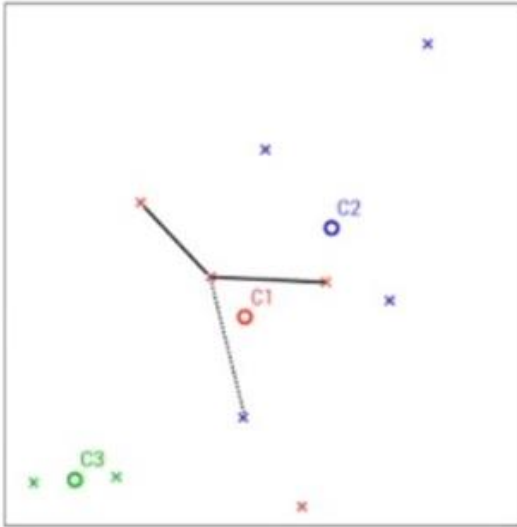
Kuruluşunuz, insanların çevrimiçi profesyonel profiller oluşturmaya olanak tanır. Önemli bir özellik, profesyonel olarak birbirine bağlı insan kümeleri oluşturma yeteneğidir. Bu kümeleri oluşturmak için ne tür bir makine öğrenimi yöntemi kullanılıyor?

- **denetimsiz makine öğrenimi**
- ikili sınıflandırma
- denetimli makine öğrenimi
- pekiştirmeli öğrenme

Q26. What is this diagram a good example of?

- K-nearest neighbor
- a decision tree
- a linear regression
- **a K-means cluster**

Note: there are centres of clusters (C0, C1, C2).



K-En Yakın Komşu(K-NN) ve K-Ortalama Kümeleme Arasındaki Fark:

K-NN, Denetimli bir makine öğrenimidir, K-ortalama ise denetimsiz bir makine öğrenimidir.

K-NN bir sınıflandırma veya regresyon makine öğrenme algoritmasıdır, K-means ise bir kümeleme makine öğrenme algoritmasıdır.

K-NN tembel bir öğrenmedir, K-Means ise istekli bir öğrenmedir. Hevesli bir öğrenme bir eğitim adımı anlamına gelen bir model uydurması vardır, ancak tembel bir öğrenmenin bir eğitim aşaması yoktur.

Tüm veriler aynı ölçeğe sahipse K-NN çok daha iyi performans gösterir, ancak bu K-ortalamlar için doğru değildir.

Q27. Random forest is modified and improved version of which earlier technique?

- aggregated trees
- boosted trees
- **bagged trees**
- stacked trees

Rastgele orman, daha önceki hangi tekniğin değiştirilmiş ve geliştirilmiş versiyonudur?

- toplu ağaçlar
- güçlendirilmiş ağaçlar
- **torbalanmış ağaçlar**
- yığılmış ağaçlar

Rastgele ormanlar veya rastgele karar ormanları, eğitim zamanında çok sayıda karar ağacı oluşturarak çalışan sınıflandırma, regresyon ve diğer görevler için bir topluluk öğrenme yöntemidir. Sınıflandırma görevleri için, rastgele ormanın çıktısı, çoğu ağaç tarafından seçilen sınıftır. Regresyon görevleri için, tek tek ağaçların ortalama veya ortalama tahmini döndürülür. Rastgele karar ormanları, karar ağaçlarının eğitim setlerine fazla uyma alışkanlıklarını düzeltir. Rastgele ormanlar genellikle karar ağaçlarından daha iyi performans gösterir.

Q28. Self-organizing maps are specialized neural network for which type of machine learning?

- **semi-supervised learning**
- supervised learning
- reinforcement learning
- unsupervised learning

Kendi kendini organize eden haritalar, hangi tür makine öğrenimi için özelleşmiş sinir ağlarıdır?

- **yarı denetimli öğrenme**
- denetimli öğrenme
- pekiştirmeli öğrenme
- denetimsiz öğrenme

Q29. Which statement about K-means clustering is true?

- **In K-means clustering, the initial centroids are sometimes randomly selected.**
- K-means clustering is often used in supervised machine learning.
- The number of clusters are always randomly selected.
- To be accurate, you want your centroids outside of the cluster.

K-arac kümelemesi ile ilgili hangi ifade doğrudur?

- **K-ortalama kümelemede, ilk merkezler bazen rastgele seçilir.**
- K-ortalama kümeleme genellikle denetimli makine öğreniminde kullanılır.
- Küme sayısı her zaman rastgele seçilir.
- Doğru olmak gerekirse, centroidlerinizin kümenin dışında olmasını istiyorsunuz.

Q30. You created machine learning system that interacts with its environment and responds to errors and rewards. What type of machine learning system is it?

- supervised learning
- semi-supervised learning
- **reinforcement learning**
- unsupervised learning

Çevresiyle etkileşime giren, hatalara ve ödüllere yanıt veren bir makine öğrenme sistemi yarattınız. Ne tür bir makine öğrenme sistemidir?

- denetimli öğrenme
- yarı denetimli öğrenme
- **pekiştirmeli öğrenme**
- denetimsiz öğrenme

Q31. Your data science team must build a binary classifier, and the number one criterion is the fastest possible scoring at deployment. It may even be deployed in real time. Which technique will produce a model that will likely be fastest for the deployment team use to new cases?

- random forest
- **logistic regression**
- KNN
- deep neural network

Veri bilimi ekibiniz bir ikili sınıflandırıcı oluşturmalıdır ve bir numaralı ölçüt, dağıtımda mümkün olan en hızlı puanlamadır. Hatta gerçek zamanlı olarak konuşlandırılabilir. Hangi teknik, dağıtım ekibinin yeni vakalarda kullanması için muhtemelen en hızlı olacak bir model üretecek?

- random forest
- **logistic regression**
- KNN
- deep neural network

Q32. Your data science team wants to use the K-nearest neighbor classification algorithm. Someone on your team wants to use a K of 25. What are the challenges of this approach?

- Higher K values will produce noisy data.
- Higher K values lower the bias but increase the variance.
- Higher K values need a larger training set.
- **Higher K values lower the variance but increase the bias.**

Veri bilimi ekibiniz K-en yakın komşu sınıflandırma algoritmasını kullanmak istiyor. Ekibinizden biri 25 K değerini kullanmak istiyor. Bu yaklaşımın zorlukları nelerdir?

- Daha yüksek K değerleri gürültülü veriler üretecektir.
- Daha yüksek K değerleri yanlılığı azaltır ancak varyansı artırır.
- Daha yüksek K değerleri daha büyük bir eğitim setine ihtiyaç duyar.
- **Daha yüksek K değerleri varyansı düşürür ancak yanlılığı artırır.**

Makine öğrenimi veri setinde iki tür gürültü olabilir: tahmin özniteliklerinde (öznitelik gürültüsü) ve hedef öznitelikte (sınıf gürültüsü). Bir veri setinde gürültünün varlığı, öğrenme algoritmalarının performansını düşüren model karmaşıklığını ve öğrenme süresini artırabilir.

Bazen algoritma yanlılığı veya AI yanlılığı olarak da adlandırılan makine öğrenimi yanlılığı, bir algoritma, makine öğrenimi sürecindeki hatalı varsayımlar nedeniyle sistematik olarak yanlış sonuçlar ürettiğinde ortaya çıkan bir olgudur.

Varyans, eğitim veri setinin farklı kısımlarını kullanırken modeldeki değişiklikleri ifade eder. Basitçe ifade etmek gerekirse, varyans, model tahminindeki değişkenliktir—verilen veri kümesine bağlı olarak ML işlevinin ne kadar ayarlayabildiği.

Q33. Your machine learning system is attempting to describe a hidden structure from unlabeled data. How would you describe this machine learning method?

- supervised learning
- **unsupervised learning**
- reinforcement learning
- semi-supervised learning

Makine öğrenimi sisteminiz, etiketlenmemiş verilerden gizli bir yapıyı tanımlamaya çalışıyor. Bu makine öğrenimi yöntemini nasıl tanımlarsınız?

- denetimli öğrenme
- **denetimsiz öğrenme**
- pekiştirmeli öğrenme
- yarı denetimsiz öğrenme

Q34. You work for a large credit card processing company that wants to create targeted promotions for its customers. The data science team created a machine learning system that groups together customers who made similar purchases, and divides those customers based on customer loyalty. How would you describe this machine learning approach?

- **It uses unsupervised learning to cluster together transactions and unsupervised learning to classify the customers.**
- It uses only unsupervised machine learning.
- It uses supervised learning to create clusters and unsupervised learning for classification.
- It uses reinforcement learning to classify the customers.

Müşterileri için hedefli promosyonlar oluşturmak isteyen büyük bir kredi kartı işleme şirketi için çalışıyorsunuz. Veri bilimi ekibi, benzer satın alma işlemleri yapan müşterileri gruplandıran ve bu müşterileri müşteri sadakatine göre bölen bir makine öğrenimi sistemi oluşturdu. Bu makine öğrenimi yaklaşımını nasıl tanımlarsınız?

- **İşlemleri bir araya getirmek için denetimsiz öğrenmeyi ve müşterileri sınıflandırmak için denetimsiz öğrenmeyi kullanır.**
- Yalnızca denetimsiz makine öğrenimini kullanır.
- Kümeler oluşturmak için denetimli öğrenmeyi ve sınıflandırma için denetimsiz öğrenmeyi kullanır.
- Müşterileri sınıflandırmak için pekiştirmeli öğrenmeyi kullanır.

Denetimsiz Öğrenme, kullanıcıların modeli denetlemesi gerekmeyen bir makine öğrenimi tekniğidir. Bunun yerine, modelin daha önce tespit edilmemiş olan kalıpları ve bilgileri keşfetmek için kendi başına çalışmasına izin verir. Esas olarak etiketlenmemiş verilerle ilgilenir.

Q35. You are using K-nearest neighbor and you have a K of 1. What are you likely to see when you train the model?

- **high variance and low bias**
- low bias and low variance
- low variance and high bias
- high bias and high variance

K-en yakın komşuyu kullanıyorsunuz ve K'niz 1'dir. Modeli eğittiğinizde muhtemelen ne göreceksiniz?

- **yüksek varyans ve düşük yanlılık**
- düşük yanlılık ve düşük varyans
- düşük varyans ve yüksek yanlılık
- yüksek yanlılık ve yüksek varyans

Bazen algoritma yanlılığı veya AI yanlılığı olarak da adlandırılan makine öğrenimi yanlılığı, bir algoritma, makine öğrenimi sürecindeki hatalı varsayımlar nedeniyle sistematik olarak yanlı sonuçlar ürettiğinde ortaya çıkan bir olgudur.

Varyans, eğitim veri setinin farklı kısımlarını kullanırken modeldeki değişiklikleri ifade eder. Basitçe ifade etmek gerekirse, varyans, model tahminindeki değişkenliktir—verilen veri kümesine bağlı olarak ML işlevinin ne kadar ayarlayabildiği.

Q36. Are data model bias and variance a challenge with unsupervised learning?

- No, data model bias and variance are only a challenge with reinforcement learning.
- **Yes, data model bias is a challenge when the machine creates clusters.**
- Yes, data model variance trains the unsupervised machine learning algorithm.
- No, data model bias and variance involve supervised learning.

Veri modeli yanlılığı ve varyansı, denetimsiz öğrenmeyle ilgili bir zorluk mudur?

- Hayır, veri modeli yanlılığı ve varyansı, pekiştirmeli öğrenmede yalnızca bir zorluktur.
- **Evet, makine kümeler oluşturduğunda veri modeli sapması bir zorluktur.**
- Evet, veri modeli varyansı denetimsiz makine öğrenimi algoritmasını eğitir.
- Hayır, veri modeli yanlılığı ve varyansı denetimli öğrenmeyi içerir.

Q37. Which choice is best for binary classification?

- K-means
- **Logistic regression**
- Linear regression
- Principal Component Analysis (PCA)

Açıklama: Lojistik regresyon, ikili sınıflandırmada lineer regresyondan çok daha iyidir, çünkü sonucu bir uçtan diğerine doğru saptırır. K-ortalama kümeleme sınıflandırma için kullanılabilir ancak çoğu senaryoda o kadar doğru değildir.

Q38. With traditional programming, the programmer typically inputs commands. With machine learning, the programmer inputs

- supervised learning
- **data**
- unsupervised learning
- algorithms

Explanation: This one is pretty straight forward and a fundamental concept.

Geleneksel programlama ile programcı tipik olarak komutları girer. Makine öğrenimi ile programcı girdileri

- denetimli öğrenme
- **veri**
- denetimsiz öğrenme
- algoritmalar

Açıklama : Bu oldukça basit ve temel bir kavram.

Q39. Why is it important for machine learning algorithms to have access to high-quality data?

- It will take too long for programmers to scrub poor data.
- If the data is high quality, the algorithms will be easier to develop.
- Low-quality data requires much more processing power than high-quality data.
- **If the data is low quality, you will get inaccurate results.**

Makine öğrenimi algoritmalarının yüksek kaliteli verilere erişmesi neden önemlidir?

- Programcıların zayıf verileri temizlemesi çok uzun sürecektir.
- Veriler yüksek kalitede ise, algoritmaların geliştirilmesi daha kolay olacaktır.
- Düşük kaliteli veriler, yüksek kaliteli verilerden çok daha fazla işlem gücü gerektirir.
- **Verilerin kalitesi düşükse, hatalı sonuçlar alırsınız.**

Q40. In K-nearest neighbor, the closer you are to neighbor, the more likely you are to

- **share common characteristics**
- be part of the root node
- have a Euclidean connection
- be part of the same cluster

K-en yakın komşuda, komşuya ne kadar yakınsanız, ... olasılığınız o kadar yüksektir.

- **ortak özellikleri paylaşma**
- kök düğümün parçası olma
- Öklid bağlantısına sahip olma
- aynı kümenin parçası olma

Q41. In the HBO show Silicon Valley, one of the characters creates a mobile application called Not Hot Dog. It works by having the user take a photograph of food with their mobile device. Then the app says whether the food is a hot dog. To create the app, the software developer uploaded hundreds of thousands of pictures of hot dogs. How would you describe this type of machine learning?

- Reinforcement machine learning
- unsupervised machine learning
- **supervised machine learning**
- semi-supervised machine learning

HBO dizisi Silicon Valley'de karakterlerden biri Not Hot Dog adlı bir mobil uygulama yaratıyor. Kullanıcının mobil cihazıyla yemeğin fotoğrafını çekmesini sağlayarak çalışır. Ardından uygulama, yemeğin sosisli sandviç olup olmadığını söylüyor. Uygulamayı oluşturmak için yazılım geliştiricisi yüz binlerce sosisli sandviç resmi yükledi. Bu tür bir makine öğrenimini nasıl tanımlarsınız?

- Güçlendirme makine öğrenimi
- denetimsiz makine öğrenimi
- **denetimli makine öğrenimi**
- yarı denetimli makine öğrenimi

Q42. You work for a large pharmaceutical company whose data science team wants to use unsupervised learning machine algorithms to help discover new drugs. What is an advantage to this approach?

- You will be able to prioritize different classes of drugs, such as antibiotics.
- You can create a training set of drugs you would like to discover.
- **The algorithms will cluster together drugs that have similar traits.**
- Human experts can create classes of drugs to help guide discovery.

Veri bilimi ekibi, yeni ilaçların keşfedilmesine yardımcı olmak için denetimsiz öğrenme makinesi algoritmalarını kullanmak isteyen büyük bir ilaç şirketinde çalışıyorsunuz. Bu yaklaşımın avantajı nedir?

- Antibiyotikler gibi farklı ilaç sınıflarına öncelik verebileceksiniz.
- Keşfetmek istediğiniz ilaçlardan oluşan bir eğitim seti oluşturabilirsiniz.
- **Algoritmalar, benzer özelliklere sahip ilaçları bir araya toplayacaktır.**
- Uzmanlar, keşfi yönlendirmeye yardımcı olacak ilaç sınıfları oluşturabilir.

Q43. In 2015, Google created a machine learning system that could beat a human in the game of Go. This extremely complex game is thought to have more gameplay possibilities than there are atoms of the universe. The first version of the system won by observing hundreds of thousands of hours of human gameplay; the second version learned how to play by getting rewards while playing against itself. How would you describe this transition to different machine learning approaches?

- **The system went from supervised learning to reinforcement learning.**
- The system evolved from supervised learning to unsupervised learning.
- The system evolved from unsupervised learning to supervised learning.
- The system evolved from reinforcement learning to unsupervised learning.

2015 yılında Google, Go oyununda bir insanı yenebilecek bir makine öğrenimi sistemi oluşturdu. Bu son derece karmaşık oyunun, evrendeki atomlardan daha fazla oynanış olanağına sahip olduğu düşünülüyor. Sistemin ilk versiyonu, yüz binlerce saatlik insan oynanışını gözlemleyerek kazandı; ikinci versiyon ise kendisine karşı oynarken ödülleri alarak oynamayı öğrendi. Farklı makine öğrenimi yaklaşımlarına geçişi nasıl tanımlarsınız?

- **Sistem denetimli öğrenmeden pekiştirmeli öğrenmeye geçti.**
- Sistem, denetimli öğrenmeden denetimsiz öğrenmeye dönüştü.
- Sistem, denetimsiz öğrenmeden denetimli öğrenmeye dönüştü.
- Sistem, pekiştirmeli öğrenmeden denetimsiz öğrenmeye dönüştü.

Q44. The security company you work for is thinking about adding machine learning algorithms to their computer network threat detection appliance. What is one advantage of using machine learning?

- **It could better protect against undiscovered threats.**
- It would very likely lower the hardware requirements.
- It would substantially shorten your development time.
- It would increase the speed of the appliance.

Çalıştığınız güvenlik şirketi, bilgisayar ağı tehdit algılama araçlarına makine öğrenimi algoritmaları eklemeyi düşünüyor. Makine öğrenimini kullanmanın avantajlarından biri nedir?

- **Keşfedilmemiş tehditlere karşı daha iyi koruma sağlayabilir.**
- Donanım gereksinimlerini büyük olasılıkla düşürür.
- Geliştirme sürenizi önemli ölçüde kısaltır.
- Cihazın hızını artıracaktır.

Q45. You work for a hospital that is tracking the community spread of a virus. The hospital created a smartwatch application that uploads body temperature data from hundreds of thousands of participants. What is the best technique to analyze the data?

- **Use reinforcement learning to reward the system when a new person participates.**
- Use unsupervised machine learning to cluster together people based on patterns the machine discovers.
- Use Supervised machine learning to sort people by demographic data.
- Use Supervised machine learning to classify people by body temperature.

Bir virüsün topluluk yayılımını izleyen bir hastane için çalışıyorsunuz. Hastane, yüz binlerce katılımcıdan vücut ısı verilerini yükleyen bir akıllı saat uygulaması oluşturdu. Verileri analiz etmek için en iyi teknik nedir?

- Yeni bir kişi katıldığında sistemi ödüllendirmek için pekiştirmeli öğrenmeyi kullanın.
- İnsanları makinenin keşfettiği kalıplara göre bir araya getirmek için denetimsiz makine öğrenimini kullanın.
- İnsanları demografik verilere göre sıralamak için Denetimli makine öğrenimini kullanın.
- **İnsanları vücut ısısına göre sınıflandırmak için Denetimli makine öğrenimini kullanın.**

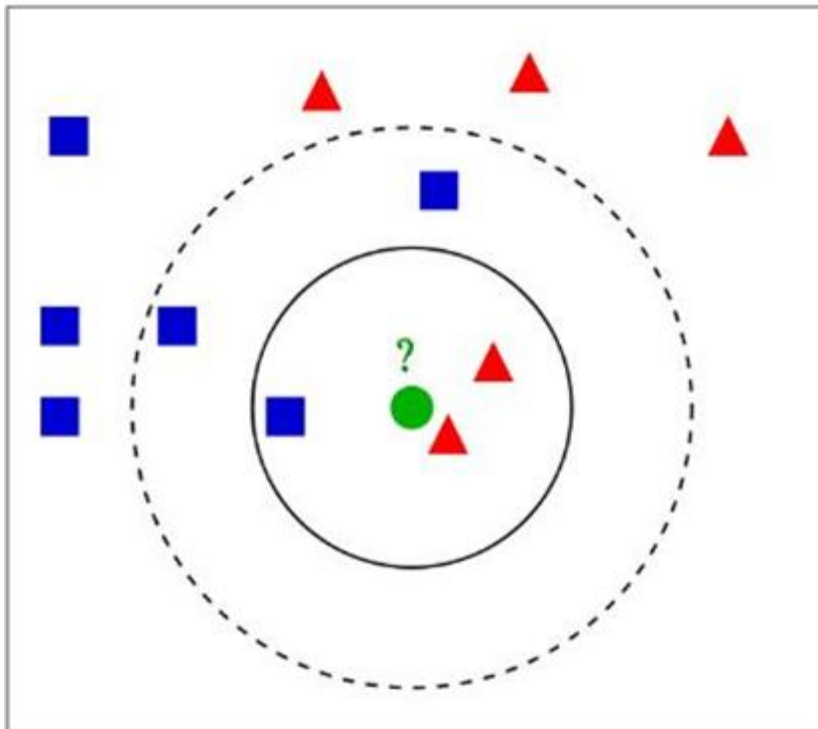
Q46. Many of the advances in machine learning have come from improved ____.

- statistics
- structured data
- availability
- **algorithms**

Makine öğrenimindeki ilerlemelerin çoğu, geliştirilmiş algoritmalarından geldi

Q47. What is this diagram a good example of?

- unsupervised learning
- complex cluster
- multiclass classification
- **k-nearest neighbour**



Q48. Naive Bayes looks at each _ predictor and creates a probability that belongs in each class.

- conditional
- multiclass
- **independent**
- binary

Naive Bayes, her _ tahmin edicisine bakar ve her sınıfa ait bir olasılık yaratır.

- koşullu
- çok sınıflı
- **bağımsız**
- ikili

Q49. Someone of your data science team recommends that you use decision trees, naive Bayes and K-nearest neighbor, all at the same time, on the same training data, and then average the results. What is this an example of?

- regression analysis
- unsupervised learning
- high -variance modeling
- **ensemble modeling**

Veri bilimi ekibinizden biri, karar ağaçlarını, Naive Bayes'i ve en yakın komşuyu aynı anda, aynı eğitim verilerinde kullanmanızı ve ardından sonuçların ortalamasını almanızı önerir. Bu neyin örneği?

- regresyon analizi
- denetimsiz öğrenme
- yüksek varyanslı modelleme
- **topluluk modelleme**

Q50. Your data science team wants to use machine learning to better filter out spam messages. The team has gathered a database of 100,000 messages that have been identified as spam or not spam. If you are using supervised machine learning, what would you call this data set?

- Machine learning algorithm
- **Training set**
- Big data test set
- Data cluster

Veri bilimi ekibiniz, istenmeyen iletileri daha iyi filtrelemek için makine öğrenimini kullanmak istiyor. Ekip, spam olarak tanımlanan veya spam olmayan 100.000 iletiden oluşan bir veritabanı topladı. Denetimli makine öğrenimi kullanıyorsanız, bu veri kümesine ne ad verirsiniz?

- Makine öğrenimi algoritması
- **Eğitim seti**
- Büyük veri test seti
- Veri kümesi

Q51. You work for a website that enables customers see all images of themselves on the internet by uploading one self-photo. Your data model uses 5 characteristics to match people to their foto: color, eye, gender, eyeglasses and facial hair. Your customers have been complaining that get tens of thousands of photos without them. What is the problem?

- You are overfitting the model to the data
- You need a smaller training set
- **You are underfitting the model to the data**
- You need a larger training set

Müşterilerin tek bir fotoğraf yükleyerek internette kendilerine ait tüm görüntüleri görmelerini sağlayan bir web sitesi için çalışıyorsunuz. Veri modeliniz, insanları fotoğraflarıyla eşleştirmek için 5 özellik kullanır: renk, göz, cinsiyet, gözlük ve sakal. Müşterileriniz on binlerce fotoğrafın onlarsız olduğundan şikayet ediyor. Sorun nedir?

- Modeli verilere fazla uyduruyorsun
- Daha küçük bir eğitim setine ihtiyacınız var
- **Modeli verilere yetersiz uyduruyorsunuz**
- Daha büyük bir eğitim setine ihtiyacınız var

Q52. Your supervisor asks you to create a machine learning system that will help your human resources department classify jobs applicants into well-defined groups. What type of system are you more likely to recommend?

- an unsupervised machine learning system that clusters together the best candidates.
- you would not recommend a machine learning system for this type of project.
- a deep learning artificial neural network that relies on petabytes of employment data.
- **a supervised machine learning system that classifies applicants into existing groups.**

Yöneticiniz, insan kaynakları departmanınızın işe başvuranları iyi tanımlanmış gruplara ayırmasına yardımcı olacak bir makine öğrenimi sistemi oluşturmanızı istiyor. Daha çok ne tür bir sistem önerirsiniz?

- en iyi adayları bir araya getiren denetimsiz bir makine öğrenimi sistemi.
- bu tür bir proje için bir makine öğrenimi sistemi önermezsiniz.
- petabaytlarca istihdam verisine dayanan derin öğrenen bir yapay sinir ağı.
- **Başvuru sahiplerini mevcut gruplara ayıran denetimli bir makine öğrenimi sistemi.**

Q53. You and your data science team have 1 TB of example data. What do you typically do with that data?

- you use it as your training set.
- **You label it big data.**
- You split it into a training set and test set.
- You use it as your test set.

Siz ve veri bilimi ekibiniz 1 TB örnek veriye sahipsiniz. Bu verilerle genellikle ne yaparsınız?

- eğitim setiniz olarak kullanıyorsunuz.
- **Büyük veri olarak etiketlersiniz.**
- Onu bir eğitim seti ve test seti olarak ayırdınız.
- Test setiniz olarak kullanıyorsunuz.

Q54. Your data science team is working on a machine learning product that can act as an artificial opponent in video games. The team is using a machine learning algorithm that focuses on rewards: If the machine does some things well, then it improves the quality of the outcome. How would you describe this type of machine learning algorithm?

- semi-supervised machine learning
- supervised machine learning
- unsupervised machine learning
- **reinforcement learning**

Veri bilimi ekibiniz, video oyunlarında yapay bir rakip olarak işlev görebilecek bir makine öğrenimi ürünü üzerinde çalışıyor. Ekip, ödüllere odaklanan bir makine öğrenimi algoritması kullanıyor: Makine bazı şeyleri iyi yaparsa, sonucun kalitesini artırır. Bu tür bir makine öğrenimi algoritmasını nasıl tanımlarsınız?

- yarı denetimli makine öğrenimi
- denetimli makine öğrenimi
- denetimsiz makine öğrenimi
- **pekiştirmeli öğrenme**

Q55. The model will be trained with data in one single batch is known as ?

- Batch learning
- Offline learning
- **Both A and B**
- None of the above

Model, verilerle eğitilecek tek bir toplu iş olarak bilinir mi?

- toplu öğrenme
- Çevrimdışı öğrenme
- **Hem A hem B**
- Yukarıdakilerin hiçbiri

Q56. Which of the following is NOT supervised learning?

- Decision Tree
- Linear Regression
- **PCA**
- Naive Bayesian

Q57. Suppose we would like to perform clustering on spatial data such as the geometrical locations of houses. We wish to produce clusters of many different sizes and shapes. Which of the following methods is the most appropriate?

- Decision Trees
- K-means clustering
- **Density-based clustering**
- Model-based clustering

Evlerin geometrik konumları gibi konumsal veriler üzerinde kümeleme yapmak istediğimizi varsayalım. Birçok farklı boyut ve şekilde kümeler üretmek istiyoruz. Aşağıdaki yöntemlerden hangisi en uygundur?

- Karar ağaçları
- K-kümeleme anlamına gelir
- **Yoğunluğa dayalı kümeleme**
- Model tabanlı kümeleme

Q58. The error function most suited for gradient descent using logistic regression is

- The entropy function.
- The squared error.
- **The cross-entropy function.**
- The number of mistakes.

Lojistik regresyon kullanan gradyan inişi için en uygun hata fonksiyonu şudur:

- Entropi işlevi.
- Kare hatası.
- **Çapraz entropi işlevi.**
- Hataların sayısı.

Q59. Compared to the variance of the Maximum Likelihood Estimate (MLE), the variance of the Maximum A Posteriori (MAP) estimate is ____

Higher

same

Lower

it could be any of the above

MAP, sonsal dağılımın en yüksek zirvesini arar, MLE ise parametreyi yalnızca verilerin olabilirlik fonksiyonuna bakarak tahmin eder.

Q60. ____ refers to a model that can neither model the training data nor generalize to new data.

- good fitting
- overfitting
- **underfitting**
- all of the above

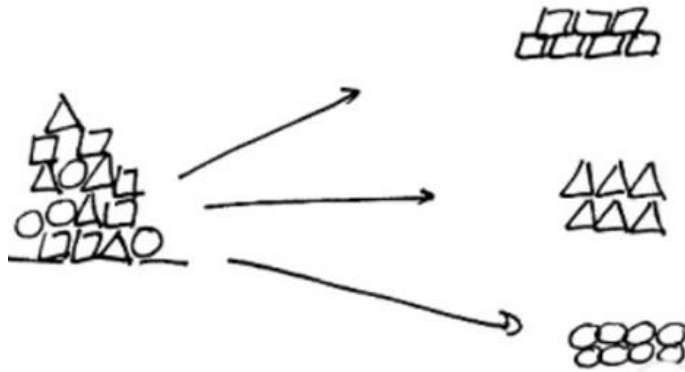
... ne eğitim verilerini modelleyebilen ne de yeni verilere genelleyemeyen bir modeli ifade eder.

- iyi uydurma
- aşırı uyum
- **yetersiz uydurma**
- Yukarıdakilerin hepsi

Q61. How would you describe this type of classification challenge?

- **This is a multiclass classification challenge.**
- This is a multi-binary classification challenge.
- This is a binary classification challenge.
- This is a reinforcement classification challenge.

Explanation: Shows data being classified into more than two categories or classes. Thus, this is a multi-class classification challenge.



Bu tür bir sınıflandırma sorununu nasıl tanımlarsınız?

- **Bu, çok sınıflı bir sınıflandırma zorluğudur.**
- Bu, çoklu ikili sınıflandırma sorunudur.
- Bu bir ikili sınıflandırma zorluğudur.
- Bu bir pekiştirme sınıflandırma zorluğudur.

Açıklama: Verilerin ikiden fazla kategoriye veya sınıfa ayrıldığını gösterir. Bu nedenle, bu çok sınıflı bir sınıflandırma zorluğudur.

Q62. What does it mean to underfit your data model?

- There is too little data in your training set.
- There is too much data in your training set.
- **There is not a lot of variance but there is a high bias.**
- Your model has low bias but high variance.

Underfitted data models usually have high bias and low variance. Overfitted data models have low bias and high variance.

Veri modelinizin altından kalkmak ne anlama geliyor?

- Eğitim setinizde çok az veri var.
- Eğitim setinizde çok fazla veri var.
- **Çok fazla varyans yok ama yüksek bir önyargı var.**
- Modelinizin önyargısı düşük ancak varyansı yüksek.

Yeterli donanıma sahip olmayan veri modelleri genellikle yüksek yanlılığa ve düşük varyansa sahiptir. Fazla donatılmış veri modelleri, düşük yanlılığa ve yüksek varyansa sahiptir.

Q63. Asian user complains that your company's facial recognition model does not properly identify their facial expressions. What should you do?

- Include Asian faces in your test data and retrain your model.
- Retrain your model with updated hyperparameter values.
- Retrain your model with smaller batch sizes.
- **Include Asian faces in your training data and retrain your model.**

The answer is self-explanatory: if Asian users are the only group of people making the complaint, then the training data should have more Asian faces.

Asyalı kullanıcı, şirketinizin yüz tanıma modelinin yüz ifadelerini doğru şekilde tanımlamadığından şikayet ediyor. Ne yapmalısın?

- Asya yüzlerini test verilerinize dahil edin ve modelinizi yeniden eğitin.
- Güncellenmiş hiperparametre değerleriyle modelinizi yeniden eğitin.
- Modelinizi daha küçük parti boyutlarıyla yeniden eğitin.
- **Eğitim verilerinize Asyalı yüzleri dahil edin ve modelinizi yeniden eğitin.**

Cevap kendiliğinden açıklayıcıdır: Şikayette bulunan tek grup Asyalı kullanıcılar, eğitim verileri daha fazla Asyalı yüze sahip olmalıdır.

Q64. You work for a website that helps match people up for lunch dates. The website boasts that it uses more than 500 predictors to find customers the perfect date, but many costumers complain that they get very few matches. What is a likely problem with your model?

- Your training set is too large.
- You are underfitting the model to the data.
- **You are overfitting the model to the data.**
- Your machine is creating inaccurate clusters.

Explanation: This question is very similar to Q49 but involves a polar opposite scenario.

that answer somewhat vague and unsettled. Small number of matchings does not necessarily implies that the model overfits, especially given 500 (!) independent variables. To me, it sounds more reasonable that the threshold (matching) criterion might be too tight, thus allowing only a small number of matching to occur. So a solution can be either softening the threshold criterion or increasing the number of candidates.

İnsanları öğle yemeği randevuları için eşleştirmeye yardımcı olan bir web sitesi için çalışıyorsunuz. Web sitesi, müşterilere mükemmel randevuyu bulmak için 500'den fazla tahmin aracı kullanmasıyla övünüyor, ancak birçok müşteri çok az eşleşme olduğundan şikayet ediyor. Modelinizle ilgili olası bir sorun nedir?

- Eğitim setiniz çok büyük.
- Modeli verilere yetersiz uyduruyorsunuz.
- **Modeli verilere fazla uyduruyorsunuz.**
- Makineniz hatalı kümeler oluşturuyor.

Açıklama: Bu soru S49'a çok benzer ancak tam tersi bir senaryo içeriyor.

bu cevap biraz belirsiz ve kararsız. Az sayıda eşleşme, özellikle 500 (!) bağımsız değişken verildiğinde, modelin fazla uyduğu anlamına gelmez. Bana göre, eşik (eşleşme) kriterinin çok sıkı olması ve dolayısıyla sadece az sayıda eşleşmenin gerçekleşmesine izin vermesi daha mantıklı geliyor. Dolayısıyla bir çözüm, eşik kriterini yumuşatmak veya aday sayısını artırmak olabilir.

Q65. (Mostly) whenever we see kernel visualizations online (or some other reference) we are actually seeing:

- **What kernels extract**
- Feature Maps
- How kernels Look

(Çoğunlukla) çevrimiçi olarak çekirdek görselleştirmelerini (veya başka bir referansı) gördüğümüzde aslında şunu görüyoruz:

- Hangi çekirdekleri çıkarır
- Özellik Haritaları
- **Çekirdekler Nasıl Görünür?**

Q66. The activations for class A, B and C before softmax were 10, 8 and 3. The different in softmax values for class A and class B would be :

- **76%**
- 88%
- 12%
- 0.0008%

A, B ve C sınıfı için softmax'tan önceki aktivasyonlar 10, 8 ve 3'tür. A sınıfı ve B sınıfı için softmax değerlerindeki fark şöyle olacaktır:

- **%76**
- %88
- %12
- %0.0008

$$\text{Softmax} := \sigma(z)_i = \frac{e^{z_i}}{\sum_j e^{z_j}} = \frac{e^{z_i}}{e^3 + e^8 + e^{10}}$$

$$\sigma(10) - \sigma(8) = \frac{e^{10} - e^8}{e^3 + e^8 + e^{10}} \approx 0.761$$

Q67. The new dataset you have just scraped seems to exhibit lots of missing values. What action will help you minimizing that problem?

- Wise fill-in of controlled random values
- Replace missing values with averaging across all samples
- Remove defective samples
- **Imputation**

Az önce kazıdığınız yeni veri kümesi birçok eksik değer sergiliyor gibi görünüyor. Hangi eylem bu sorunu en aza indirmenize yardımcı olacak?

- Kontrollü rastgele değerlerin akıllıca doldurulması
- Eksik değerleri tüm örneklerde ortalama olarak değiştirin
- Arızalı numuneleri çıkarın
- **Atama**

Q68. Which of the following methods can use either as an unsupervised learning or as a dimensionality reduction technique?

- SVM
- **PCA**
- LDA
- TSNE

Aşağıdaki yöntemlerden hangisi denetimsiz öğrenme veya boyut azaltma tekniği olarak kullanılabilir?

Açıklama: Temel Bileşen Analizi (Principal Component Analysis - PCA), çeşitli uygulamalarda en sık kullanılan denetimsiz makine öğrenimi algoritmalarından biridir: keşifsel veri analizi, boyutsallık azaltma, bilgi sıkıştırma, veri gürültüsüzleştirme ve çok daha fazlası!

Q69. What is the main motivation for using activation functions in ANN?

- **Capturing complex non-linear patterns**
- Transforming continuous values into "ON" (1) or "OFF" (0) values
- Help avoiding the vanishing/exploding gradient problem
- Their ability to activate each neurons individually.

YSA'da (Yapay Sinir Ağları) aktivasyon fonksiyonlarını kullanmanın ana motivasyonu nedir?

- **Karmaşık doğrusal olmayan desenleri yakalama**
- Sürekli değerlerin "AÇIK" (1) veya "KAPALI" (0) değerlerine dönüştürülmesi
- Kaybolan/patlayan gradyan probleminde kaçınmaya yardımcı olun
- Her nöronu ayrı ayrı aktive etme yetenekleri.

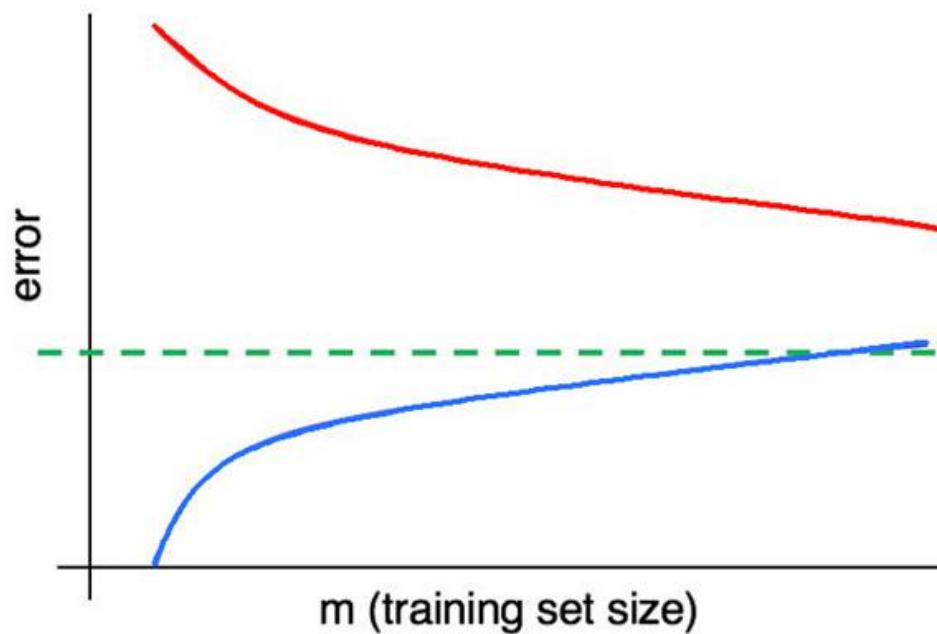
Q70. Which loss function would fit best in a categorical (discrete) supervised learning ?

- Kullback-leibler (KL) loss
- **Binary Crossentropy**
- Mean Squared Error (MSE)
- Any L2 loss

Kategorik (ayrık) denetimli öğrenmeye en uygun kayıp işlevi hangisidir?

- Kullback-leibler (KL) kaybı
- **İkili Çapraz Entropi**
- Ortalama Kare Hatası (MSE)
- Herhangi bir L2 kaybı

Q71. What is the correct option?



no.	Red	Blue	Green
1.	Validation error	Training error	Test error
2.	Training error	Test error	Validation error
3.	Optimal error	Validation error	Test error
4.	Validation error	Training error	Optimal error

Q72. You create a decision tree to show whether someone decides to go to the beach. There are three factors in this decision: rainy, overcast, and sunny. What are these three factors called?

- Tree nodes
- **Predictors**
- Root nodes
- Deciders

Explain: These nodes decide whether the someone decides to go to beach or not, for example if its rainy people will mostly refrain from going to beach

Birinin sahile gitmeye karar verip vermediğini göstermek için bir karar ağacı yaratırsınız. Bu kararda üç faktör vardır: yağmurlu, bulutlu ve güneşli. Bu üç faktöre ne ad verilir?

- Ağaç düğümleri
- **Tahmin ediciler**
- Kök düğümler
- Karar vericiler

Açıklama: Bu düğümler, birinin kumsala gitmeye karar verip vermeyeceğine karar verir, örneğin, yağmurlu insanlar çoğunlukla kumsala gitmekten kaçınacaklardır.

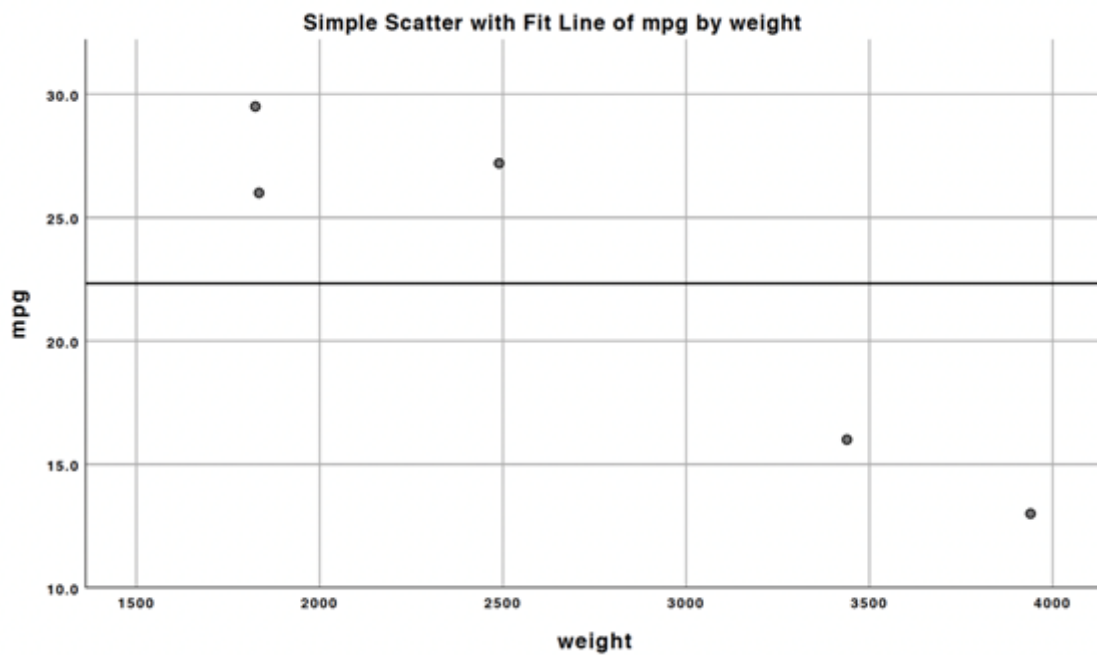
Q73. You need to quickly label thousands of images to train a model. What should you do?

- Set up a cluster of machines to label the images
- Create a subset of the images and label then yourself
- **Use naive Bayes to automatically generate labels.**
- Hire people to manually label the images

Bir modeli eğitmek için binlerce görüntüyü hızla etiketlemeniz gerekir. Ne yapmalısınız?

- Görüntüleri etiketlemek için bir makine kümesi kurun
- Görüntülerin bir alt kümesini oluşturun ve ardından kendinizi etiketleyin
- **Etiketleri otomatik olarak oluşturmak için Naive Bayes kullanın.**
- Görüntüleri manuel olarak etiketlemek için insanları işe alın

Q74. The fit line and data in the figure exhibits which pattern?



- low bias, high variance
- high bias, low variance
- high bias, high variance
- **low bias, low variance**

Explain: Since the data is accurately classified and is neither overfitting or underfitting the dataset

Şekildeki uyum çizgisi ve veriler hangi kalıbı göstermektedir?

- düşük önyargı, yüksek varyans
- yüksek önyargı, düşük varyans
- yüksek önyargı, yüksek varyans
- **düşük önyargı, düşük varyans**

Açıklama: Veriler doğru bir şekilde sınıflandırıldığından ve veri kümesine ne fazla uyuyor ne de yetersiz uyuyor

Q75. Many of the advances in machine learning have come from improved?

- structured data
- **algorithms**
- time
- computer scientists

Makine öğrenimindeki ilerlemelerin çoğu, iyileştirmelerden mi geldi?

- yapılandırılmış veri
- **algoritmalar**
- zaman
- Bilgisayar bilimcileri

Q76. You need to select a machine learning process to run a distributed neural network on a mobile application. Which would you choose?

- Scikit-learn
- PyTorch
- **Tensowflow Lite**
- Tensorflow

Bir mobil uygulamada dağıtılmış bir sinir ağını çalıştırmak için bir makine öğrenimi süreci seçmeniz gerekir. Hangi seçerdiniz?

- Scikit-öğren
- PyTorch
- **Tensowflow Lite**
- tensör akışı

Q77. Which choice is the best example of labeled data?

- **a spreadsheet**
- 20,000 recorded voicemail messages
- 100,000 images of automobiles
- hundreds of gigabytes of audio files

Etiketlenmiş verilere en iyi örnek hangi seçimdir?

- **bir elektronik tablo**
- 20.000 kayıtlı sesli posta mesajı
- 100.000 otomobil resmi
- yüzlerce gigabayt ses dosyası

Q78. In statistics, what is defined as the probability of a hypothesis test of finding an effect - if there is an effect to be found?

- Confidence
- Alpha
- **Power**
- Significance

İstatistikte, bir etki bulma hipotez testinin olasılığı olarak tanımlanır - eğer bulunacak bir etki varsa?

- Kendinden emin
- Alfa
- **Güç**
- Önem

Q79. You want to create a machine learning algorithm to identify food recipes on the web. To do this, you create an algorithm that looks at different conditional probabilities. So if the post includes the word flour, it has a slightly stronger probability of being a recipe. If it contains both flour and sugar, it even more likely a recipe. What type of algorithm are you using?

- **Naive Bayes classifier**
- K-nearest neighbor
- Multiclass classification
- Decision tree

Web'deki yemek tariflerini tanımlamak için bir makine öğrenimi algoritması oluşturmak istiyorsunuz. Bunu yapmak için farklı koşullu olasılıklara bakan bir algoritma oluşturursunuz. Yani gönderi un kelimesini içeriyorsa, tarif olma olasılığı biraz daha yüksek. Hem un hem de şeker içeriyorsa, bir tarif olması daha olasıdır. Ne tür bir algoritma kullanıyorsunuz?

- **Naive Bayes sınıflandırıcı**
- K-en yakın komşu
- Çok sınıflı sınıflandırma
- Karar ağacı

Q80. What is lazy learning?

- When the machine learning algorithms do most of the programming
- When you don't do any data scrubbing
- When the learning happens continuously
- **When you run your computation in one big instance at the beginning**

Tembel öğrenme nedir?

- Programlamanın çoğunu makine öğrenimi algoritmaları yaptığında
- Herhangi bir veri temizleme yapmadığınızda
- Öğrenme sürekli gerçekleştiğinde
- **Hesaplamanızı başlangıçta büyük bir örnekte çalıştırdığınızda**

Açıklama: Makine öğreniminde tembel öğrenme, sistemin sorguları almadan önce eğitim verilerini genelleştirmeye çalıştığı istekli öğrenmenin aksine, eğitim verilerinin genelleştirilmesinin teorik olarak sisteme bir sorgu yapılana kadar ertelendiği bir öğrenme yöntemidir. .

Q81. What is Q-learning reinforcement learning?

- Supervised machine learning with rewards
- A type of unsupervised learning that relies heavily on a well-established model
- A type of reinforcement learning where accuracy degrades over time
- **A type of reinforcement learning that focuses on rewards**

Explanation:Q-learning is a model-free reinforcement learning algorithm.Q-learning is a values-based learning algorithm. Value based algorithms updates the value function based on an equation(particularly Bellman equation). Reference

Q-öğrenme pekiştirmeli öğrenme nedir?

- Ödüllü denetimli makine öğrenimi
- Büyük ölçüde köklü bir modele dayanan bir tür denetimsiz öğrenme
- Doğruluğun zamanla azaldığı bir tür pekiştirmeli öğrenme
- **Ödüllere odaklanan bir tür pekiştirmeli öğrenme**

Açıklama:Q-learning, modelsiz bir pekiştirmeli öğrenme algoritmasıdır.Q-learning, değerlere dayalı bir öğrenme algoritmasıdır. Değer tabanlı algoritmalar, bir denkleme (özellikle Bellman denklemi) dayalı olarak değer fonksiyonunu günceller.

Q-learning, belirli bir durumdaki bir eylemin değerini öğrenmek için modelsiz bir pekiştirmeli öğrenme algoritmasıdır. Bir çevre modeli gerektirmez (dolayısıyla "modelsiz") ve stokastik geçişler ve ödüllerle ilgili sorunları uyarlamalar gerektirmeden halledebilir.

Q82. The data in your model has low bias and low variance. How would you expect the data points to be grouped together on the diagram?

- **They would be grouped tightly together in the predicted outcome.**
- They would be grouped tightly together but far from the predicted.
- They would be scattered around the predict outcome.
- They would be scattered far away from the predicted outcome.

Modelinizdeki veriler düşük sapma ve düşük varyansa sahiptir. Veri noktalarının diyagramda birlikte gruplanmasını nasıl beklersiniz?

- **Öngörülen sonuçta birlikte sıkı bir şekilde gruplandırılacaklardı.**
- Birbirlerine sıkı sıkıya ama tahmin edilenden uzak bir şekilde gruplandırılacaklardı.
- Tahmin sonucunun etrafına dağılmış olacaklardı.
- Tahmin edilen sonuçtan çok uzaklara dağılmış olacaklardı.

Q83. Your machine learning system is using labeled examples to try to predict future data, compare that data to the predicted result, and then the model. What is the best description of this machine learning method?

- Unsupervised learning
- Semi-supervised learning
- **Supervised learning**
- Semi-reinforcement learning

Makine öğrenimi sisteminiz, gelecekteki verileri tahmin etmeye çalışmak, bu verileri tahmin edilen sonuçla ve ardından modelle karşılaştırmak için etiketli örnekler kullanıyor. Bu makine öğrenimi yönteminin en iyi açıklaması nedir?

- Denetimsiz öğrenme
- Yarı denetimli öğrenme
- **Denetimli öğrenme**
- Yarı pekiştirmeli öğrenme

Q84. In the 1983 movie WarGames, the computer learns how to master the game of chess by playing against itself. What machine learning method was the computer using?

- Binary learning
- Supervised learning
- Unsupervised learning
- **Reinforcement learning**

1983 yapımı WarGames filminde bilgisayar, kendisine karşı oynayarak satranç oyununda ustalaşmayı öğrenir. Bilgisayar hangi makine öğrenimi yöntemini kullanıyordu?

- İkili öğrenme
- Denetimli öğrenme
- Denetimsiz öğrenme
- **Pekiştirmeli öğrenme**

Q85. You are working with your machine learning algorithm on something called class predictor probability. What algorithm are you most likely using?

- Multiclass binary classification
- **Naive bayes**
- Unsupervised classification
- Decision tree analysis

You could use a naïve Bayes algorithm, to differentiate three classes of dog breeds — terrier, hound, and sport dogs. Each class has three predictors — hair length, height, and weight. The algorithm does something called class predictor probability.

Makine öğrenimi algoritmanızla, sınıf tahmin etme olasılığı denen bir şey üzerinde çalışıyorsunuz. Büyük olasılıkla hangi algoritmayı kullanıyorsunuz?

- Çok sınıflı ikili sınıflandırma
- **Naive bayes**
- Denetimsiz sınıflandırma
- Karar ağacı analizi

Terrier, tazı ve spor köpekleri olmak üzere üç köpek ırkı sınıfını ayırt etmek için saf bir Bayes algoritması kullanabilirsiniz. Her sınıfın üç belirleyicisi vardır - saç uzunluğu, boy ve ağırlık. Algoritma, sınıf tahmin etme olasılığı adı verilen bir şey yapar.

Q86. What is one of the most effective way to correct for underfitting your model to the data?

- Create training clusters
- Remove predictors
- Use reinforcement learning
- **Add more predictors**

Modelinizin verilere uymamasını düzeltmenin en etkili yollarından biri nedir?

- Eğitim kümeleri oluşturun
- Tahmin edicileri kaldır
- Takviyeli öğrenmeyi kullanın
- **Daha fazla tahminci ekleyin**

Q87. Your data science team is often criticized for creating reports that are boring or too obvious. What could you do to help improve the team?

- **Suggest that the team is probably underfitting the model to the data.**
- Suggest that unsupervised learning will lead to more interesting results.
- Make sure that they are picking the correct machine learning algorithms.
- Encourage the team to ask more interesting questions.

Veri bilimi ekibiniz genellikle sıkıcı veya çok açık raporlar oluşturduğu için eleştirilir. Takımın gelişmesine yardımcı olmak için ne yapabilirsin?

- **Takımın muhtemelen modeli verilere yetersiz uydurduğunu önerin.**
- Denetimsiz öğrenmenin daha ilginç sonuçlara yol açacağını önerin.
- Doğru makine öğrenimi algoritmalarını seçtiklerinden emin olun.
- Ekibi daha ilginç sorular sormaya teşvik edin.

Q88. What is the difference between unstructured and structured data?

- Unstructured data is always text.
- Unstructured data is much easier to store.
- **Structured data has clearly defined data types.**
- Structured data is much more popular.

Yapılandırılmamış ve yapılandırılmış veriler arasındaki fark nedir?

- Yapılandırılmamış veriler her zaman metindir.
- Yapılandırılmamış verilerin saklanması çok daha kolaydır.
- **Yapılandırılmış veriler, açıkça tanımlanmış veri türlerine sahiptir.**
- Yapılandırılmış veriler çok daha popüler.

Q89. You work for a startup that is trying to develop a software tool that will scan the internet for pictures of people using specific tools. The chief executive is very interested in using machine learning algorithms. What would you recommend as the best place to start?

- Using an unsupervised machine learning algorithm to cluster together all the photographs.
- Create a data lake with an unsupervised machine learning algorithm.
- Use a combination of unsupervised and supervised machine learning to create machine-defined data clusters.
- **Use supervised machine learning to classify photographs based on a predetermined training set.**

Belirli araçları kullanan kişilerin resimleri için interneti tarayacak bir yazılım aracı geliştirmeye çalışan bir girişim için çalışıyorsunuz. CEO, makine öğrenimi algoritmalarını kullanmakla çok ilgileniyor. Başlamak için en iyi yer olarak ne önerirsiniz?

- Tüm fotoğrafları bir araya getirmek için denetimsiz bir makine öğrenimi algoritması kullanmak.
- Denetimsiz bir makine öğrenimi algoritmasıyla bir veri gölü oluşturun.
- Makine tanımlı veri kümeleri oluşturmak için denetimsiz ve denetimli makine öğreniminin bir kombinasyonunu kullanın.
- **Fotoğrafları önceden belirlenmiş bir eğitim setine göre sınıflandırmak için denetimli makine öğrenimini kullanın.**

Q90. In supervised machine learning, data scientist often have the challenge of balancing between underfitting or overfitting their data model. They often have to adjust the training set to make better predictions. What is this balance called?

- the under/over challenge
- balance between clustering classification
- **bias-variance trade-off**
- the multiclass training set challenge

Denetimli makine öğreniminde, veri bilimcisi genellikle veri modellerini eksik veya fazla uydurma arasında denge kurma zorluğuyla karşılaşır. Daha iyi tahminler yapmak için genellikle eğitim setini ayarlamak zorunda kalırlar. Bu dengeye ne denir?

- alt/üst meydan okuma
- kümeleme sınıflandırması arasındaki denge
- **önyargı-varyans değiş tokuşu**
- çok sınıflı eğitim seti mücadelesi

Q91. What is conditional probability?

- **the probability that doing one thing has an impact on another thing**
- the probability that certain conditions are met
- the probability that, based on certain conditions, something will always be incorrect
- the probability of something being the correct answer

Koşullu olasılık nedir?

- **bir şeyi yapmanın başka bir şeyi etkileme olasılığı**
- belirli koşulların karşılanma olasılığı
- Belirli koşullara bağlı olarak bir şeyin her zaman yanlış olma olasılığı
- bir şeyin doğru cevap olma olasılığı

Q92. K-means clustering is what type of machine learning algorithm?

- reinforcement
- supervised
- **unsupervised**
- classification

Q93. What is ensemble modeling?

- when you create an ensemble of your training and test data set
- when you create an ensemble of different servers to run the algorithms
- when you find the one best algorithm for your ensemble
- **when you use several ensembles of machine learning algorithms**

Topluluk modellemesi nedir?

- eğitim ve test veri setinizin bir grubunu oluşturduğunuzda
- algoritmaları çalıştırmak için farklı sunuculardan oluşan bir topluluk oluşturduğunuzda
- topluluğunuz için en iyi algoritmayı bulduğunuzda
- **birkaç makine öğrenimi algoritması topluluğu kullandığınızda**

Toplu modelleme, ya birçok farklı modelleme algoritması kullanarak ya da farklı eğitim veri setleri kullanarak bir sonucu tahmin etmek için çok çeşitli modellerin oluşturulduğu bir süreçtir. Topluluk modeli daha sonra her bir temel modelin tahminini toplar ve görünmeyen veriler için bir kez nihai tahminle sonuçlanır.

Q94. What is the best definition for bias in your data model?

- Bias is when your predicted values are scattered.
- **Bias is the gap between your predicted value and the outcome.**
- Bias is when your data is wrong for different reasons.
- Bias is when your values are always off by the same percentage.

Veri modelinizde yanlışlık için en iyi tanım nedir?

- Yanlılık, tahmin edilen değerlerinizin dağıldığı zamandır.
- **Yanlılık, tahmin edilen değeriniz ile sonuç arasındaki boşluktur.**
- Yanlılık, verilerinizin farklı nedenlerle yanlış olduğu zamandır.
- Yanlılık, değerlerinizin her zaman aynı yüzdeyle kapalı olduğu zamandır.

Kaynakça

- 1- <https://github.com/Ebazhanov/linkedin-skill-assessments-quizzes/blob/main/machine-learning/machine-learning-quiz.md>