

2 Neden Makine Öğrenmesi?

Makine öğrenimi önemlidir çünkü bilgisayarların verilerden öğrenmesine ve açıkça programlanmadan belirli görevlerdeki performanslarını artırmasına olanak tanır. Verilerden öğrenme ve yeni durumlara uyum sağlama yeteneği, makine öğrenimini özellikle büyük miktarda veri, karmaşık karar verme ve dinamik ortamlar içeren görevler için yararlı kılar.

Makine öğreniminin kullanıldığı bazı spesifik alanlar şunlardır:

- Tahmine dayalı modelleme: Makine öğrenimi, işletmelerin daha iyi kararlar almasına yardımcı olabilecek tahmine dayalı modeller oluşturmak için kullanılabilir. Örneğin, makine öğrenimi, hangi müşterilerin belirli bir ürünü satın alma olasılığının yüksek olduğunu veya hangi hastaların belirli bir hastalığa yakalanma olasılığının yüksek olduğunu tahmin etmek için kullanılabilir.
- Doğal dil işleme: Makine öğrenimi, insan dilini anlayabilen ve yorumlayabilen sistemler oluşturmak için kullanılır. Bu, ses tanıma, sohbet robotları ve dil çevirisi gibi uygulamalar için önemlidir.
- Bilgisayarla görme: Makine öğrenimi, görüntüleri ve videoları tanıyabilen ve yorumlayabilen sistemler oluşturmak için kullanılır. Bu, sürücüsüz arabalar, gözetim sistemleri ve tıbbi görüntüleme gibi uygulamalar için önemlidir.
- Dolandırıcılık tespiti: Makine öğrenimi, finansal işlemlerde, çevrimiçi reklamlarda ve diğer alanlardaki dolandırıcılık davranışlarını tespit etmek için kullanılabilir.
- Öneri sistemleri: Makine öğrenimi, kullanıcılara geçmiş davranışlarına ve tercihlerine göre ürün, hizmet veya içerik öneren öneri sistemleri oluşturmak için kullanılabilir.

Makine öğrenimi: Örnek verileri veya geçmiş deneyimleri kullanarak bir performans kriterini optimize etmek için bilgisayarları programlamaktır. İyi olasılıklı modeller oluşturarak bir veri kütesinden faydalı bilgilerin otomatik olarak çıkarılması. Genel bir teorinin olmadığı durumlarda çok fazla veriye sahip alanlar için idealdir.

Makine Öğrenimi çalışması: Deneyimle bazı görevlerde performanslarını artıran algoritmaların geliştirilmesidir. Bir çıkarım süreci, model uydurma veya örneklerden öğrenme yoluyla, verilerden teoriyi otomatik olarak öğrenmek.

- Örnekler dışında bazı görevler iyi tanımlanamaz (örneğin yüzlerin veya insanların tanınması).
- Büyük miktarda verinin gizli ilişkileri ve korelasyonları olabilir. Bunları yalnızca otomatik yaklaşımlar tespit edebilir.

- Belirli bir sorun / görev hakkındaki bilgi miktarı, insanlar tarafından açık kodlama için çok büyük olabilir (örneğin, tıbbi teşhislerde)
- Ortamlar zamanla değişir ve sürekli olarak yeni bilgiler keşfedilir. Sistemlerin "elle" sürekli olarak yeniden tasarlanması zor olabilir.

Makine öğrenmesinde:

- Öğrenme Görevi: Ne öğrenmek ya da tahmin etmek istiyoruz?
- Veriler ve varsayımlar: Elimizde hangi veriler var? Kaliteleri nedir? Verilen problem hakkında ne varsayabiliriz?
- Temsil: Sınıflandırılacak örneklerin uygun bir temsili nedir? • Yöntem ve Tahmin: Olası hipotezler var mı? Tahminlerimizi verilen sonuçlara göre ayarlayabilir miyiz?
- Değerlendirme: Yöntem ne kadar iyi performans gösteriyor? Başka bir yaklaşım / model daha iyi performans gösterebilir mi?
- Sınıflandırma: Bir öge sınıfının tahmini.
- Öngörü: Bir parametre değerinin tahmini.
- Karakterizasyon: Öge gruplarını tanımlayan hipotezler bulun.
- Kümeleme: (Atanmamış) veri kümesinin ortak özelliklere sahip kümelere bölünmesi. (Denetimsiz öğrenme)

Örneğin, bir dizi kedi ve köpek resmimiz var. Yapmak istediğimiz şey onları bir grup kedi ve köpek olarak sınıflandırmak. Bunu yapmak için farklı hayvan özelliklerini bulmamız gerekiyor, örneğin:

- Her hayvanın kaç gözü vardır?
- Her hayvanın göz rengi nedir?
- Her bir hayvanın boyu kaçtır?
- Her bir hayvanın uzunluğu nedir?
- Her bir hayvanın bacak uzunlukları nedir?
- Her bir hayvanın ağırlığı nedir?
- Her hayvan genellikle ne yer?

Bu soruların her birinin cevabı için bir vektör oluşturulur. Ardından, bir dizi kural uygulanır: Boy > 30cm ve ağırlık > 2Kg ise, o zaman bir köpek olabilir. Her veri noktası için böyle bir dizi kurallar dizisi oluşturulur. Ayrıca, if, else if, else ifadelerinden oluşan bir karar ağacı oluşturulur ve kategorilerden birine girip girmediğini kontrol edilir. Makine öğreniminin yaptığı şey, verileri farklı algoritmalarla işlemektir ve bize bunun bir kedi mi yoksa köpek mi

olduğunu belirlemede hangi özelliğin daha önemli ve etkin olduğunu söyler. O halde sonuca etkisi fazla olan parametrelerin belirlenmesi çok daha önemsenmelidir. Bu nedenle, birçok kural seti uygulamak yerine, bunu iki veya üç özelliğe göre basitleştirebiliriz ve sonuç olarak bu bize daha yüksek bir doğruluk sağlar.

Genellikle, makine öğrenimi yöntemleri iki aşamaya ayrılır:

1) Eğitim modeli: Verilerin bir koleksiyonundan öğrenilir.

2) Uygulama modeli: Yeni test verileri hakkında kararlar almak için kullanılır.

Makine öğrenimi türlerinden bazıları şunlardır:

- Eğitim verilerinin doğru yanıtlarla etiketlendiği denetimli öğrenim. En yaygın iki denetimli öğrenme türü, sınıflandırma ve regresyondur.
- Analiz etmek ve keşfetmek istediğimiz kalıpları, etiketlenmemiş verilerden oluşan bir koleksiyondan öğrenen denetimsiz öğrenme. En önemli iki örnek, boyut küçültme ve kümelenmedir.
- Robot veya kontrolör gibi bir temsilcinin geçmişteki eylemlerin sonuçlarına dayalı olarak uygun eylemleri öğrenmeye çalıştığı pekiştirmeli öğrenme.
- Eğitim verilerinin yalnızca bir alt kümesinin etiketlendiği yarı denetimli öğrenme.
- Mali piyasalarda olduğu gibi zaman serisi tahmini
- Fabrikalarda ve gözetimde arıza tespiti için kullanılanlar gibi anormallik tespiti.
- Verilerin elde edilmesinin pahalı olduğu aktif öğrenme

Bu nedenle bir algoritmanın hangi eğitim verilerinden elde edileceğini ve diğerlerini belirlenmesi gerekir.

Makine öğrenimi modeli nedir?

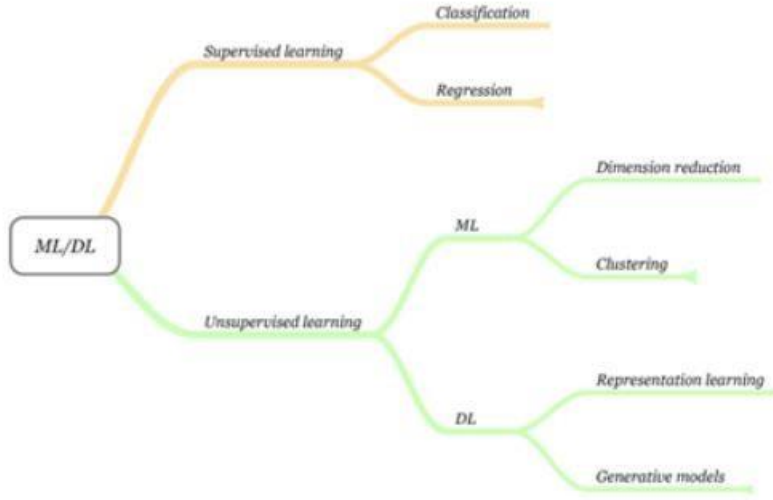
Bir makine öğrenimi modeli, makine öğrenimiyle ilgili görevleri işleme koyan bir soru ya da yanıtlama sistemidir. İş sonuçlarını iyileştirmek için değerli içgörüler toplamak, etkin olan parametrelere belirlemek için verileri kullanmayı amaçlar.

Disiplinler:

Mühendislikler: Bilgisayar, Elektronik, Makine

Matematik: Olasılık, istatistik, Nümerik analiz, İşaret işleme, Uygulamalı matematik, Stokastik, Algoritma

Antropoloji, Zooloji, Biyoloji, Bakteriyoloji, Genetik, Nörobiyoloji, Paleontoloji Uzay bilimleri



Deneyim Performans ölçüsü:

Makine öğrenimi, bir çocuğun büyümesine benzer şekilde davranır. Bir çocuk büyüdükçe, görevi yerine getirme deneyimi artar ve bu da performans ölçüsü ile sonuçlanır. Örneğin bir çocuğa “şekil ayırma bloğu” oyuncacı verilirse, (Artık hepimiz bu oyuncakta farklı şekil ve şekillerde deliklerimiz olduğunu biliyoruz). Bu durumda, görev, bir şekil için uygun bir şekil deliği bulmaktır. Daha sonra çocuk şekli gözlemler ve şekilli bir deliğe yerleştirmeye çalışır. Diyelim ki bu oyuncakın üç şekli var: bir daire, bir üçgen ve bir kare. Şekilli bir delik bulmaya yönelik ilk denemesinde performans ölçüsü 1/3'tür, bu da çocuğun 3 doğru şekil deliğinden 1'ini bulduğu anlamına gelir.

İkincisi, çocuk başka bir zaman dener ve bu görevde biraz tecrübeli olduğunu fark eder. Kazanılan deneyime bakıldığında çocuk bu görevi başka bir zaman dener ve performansı ölçerken 2/3 olduğu ortaya çıkar. Bu görevi 100 kez tekrarladıktan sonra bebek şimdi hangi şeklin hangi şekil deliğine gireceğini anladı. Böylece deneyimi arttı, performansı da arttı ve sonra bu oyuncakta yapılan deneme sayısı arttıkça fark edilirdi, performans da artar ve bu da daha yüksek doğruluk sağlar. Bu tür bir uygulama, makine öğrenimine benzer. Bir makinenin yaptığı şey, bir görevi alır, onu yürütür ve performansı ölçer. Artık bir makinenin çok sayıda verisi vardır, bu nedenle bu verileri işledikçe deneyimi zamanla artar ve daha yüksek bir performans ölçüsü ile sonuçlanır. Dolayısıyla, tüm verileri gözden geçirdikten sonra, makine öğrenimi modelimizin doğruluğu artar, bu da modelimiz tarafından yapılan tahminlerin çok doğru olacağı anlamına gelir.

Bu çalışmada, hesaplama yapıları da dikkate alınacaktır:

- Fonksiyonlar
- Mantıksal tasarım kuralları, sonlu durum makineleri
- Dilbilgisi
- Problem çözme mantıkları

- İstatiksel analiz metotları
- Stokastik süreçler teorisi
- Uygulamalı matematik
- Kaos ve Sapma analisi
- Akıl Oyunları
- Belirsizlik • Hata Kaynakları

Büyük Veri:

Yapay zekâ teknolojisinin altında büyük veri (big data) yatıyor. Verinin gerçek zamanlı işlenebilmesi, bu datanın farklı kanallardan gelmesi, kendi kendine öğrenen teknolojilere sahip uygun ürünlere ulaşabilmesini sağlamak önemlidir. Artık kural bazlı statik teknolojiler yetmiyor. Yapay zekâyı bu yüzden teknoloji açısından değer yaratacak şekilde kullanıyoruz.

Gelecekte teknoloji departmanlarının yarısından fazlası veri bilimcilerden (Data Scientist) oluşacaktır. Beklenen önemli değişimlerden biri fiziksel dünya ile dijital dünyanın artık iç içe geçmişliği örneğin süt bittiğine otomatik sipariş geçen buzdolaplarından insanların çalışmadığı mağazalara uzanan bir dönüşümden bahsediyoruz. Bu dönüşümün ortasında da yapay zekâ teknolojileri oturuyor.

Şurası çok önemli; çağımızın insanı kendisine özel bir dokunuş istiyor. Yani bir grubun içerisinde addedilerek, o grubun bir parçası olan ortak bir kimlikmiş gibi addedilmek istemiyor. Mutlaka ona özel bir dokunuş” dedim. Tek tek herkese ayrı hitap etmeniz gerekiyor. Mesela ben şahsıma özelleşmemiş hiçbir maili cevaplamam. İsmen hitap edilen, hukukumun gereği olan, karşılığı olan, üslubu olmayan standart hiçbir maili cevaplamam. Ben yokum ki orada. Bir topluluğa atılmış. Kim cevabını veriyorsa versin.

İnsanoğlunun anne karnından itibaren hareket ettiğini ve bunun hayatın sonuna kadar değişmedi. Yapılan bir araştırmaya göre sayfa yüklenme süresindeki her 1 saniyelik hızlanma internet satışlarını yüzde 5 arttırıyor.

Bilgisayar kontrolündeki bir sistemin ya da cihazın faaliyetlerini hafızasındaki verileri kullanarak insan zekasına benzer şekilde yerine getirme yeteneği kazandırılması, veri yığınının öğrenen zekanın geliştirilmesi ile mümkün olabilir. Diğer bir anlatımla, **kendi kendine öğrenen matematiksel modeller ve algoritmalar ile insandan bağımsız davranış geliştirmesi gerekmektedir.**

Gezgin algılayıcıların, sistemlerin ve makinelerin artması, sosyal ağlara gezgin erişimin yaygınlaşması, çeşitli takip (sensörler, barkodlar, karekodlar, RFID sistemleri... vs.), nesnelerin interneti (IOT) ve otomasyon teknolojilerinin gelişmesi, iletişim teknolojilerinin ulaşılabilirliğinin artması, başta ticari işlemler olmak üzere pek çok iş kolunun elektronik ortama taşınması ile birlikte hem üretilen verinin çeşitliliği hem de toplanma hızı ve miktarı da ciddi oranlarda artmıştır. Bu artış üstel olarak devam etmektedir.

Öte yandan cihazlara takılan sim kartlar, algılayıcılar, elektronik ölçerler, bilgisayar sistemleri ve yazılımları sayesinde, cihazların uzaktan izlenmesini, yönetilmesini ve internet ağı ile birbiriyle iletişim kurabilmesini sağlayan bir teknoloji olan Makineler Arası İletişim (M2M) ve Nesnelerin İnterneti, hem bireylerin hem de şirketlerin hayatında geniş bir kullanım alanı bulmaktadır.

M2M sistemleri, günümüzde neredeyse herhangi bir donanımın çeşitli uygulamalar veya cihazlarla birbirine kolayca bağlanabildiği Nesnelerin İnterneti (IoT), Her şeyin İnterneti (IoE), Nesnelerin Ağı (WoT) ve Her şeyin Ağı (WoE) gibi ortamlara evrilmiştir. Akıllı ortamların meydana geldiği bu sistemlerde, muazzam bir veri hacmi üretilir ve üstelik bu verilerin çoğu yapılandırılmamıştır. Resim, ses, metin, video gibi pek çok türde olabilen ve ağlar üzerinden aktarılan bu veriler bulut ortamlarda da saklanmaya başlamıştır. Bu verilerle ilgili bir başka husus ise sosyal medya verileri gibi insan kaynaklı veriler başta olmak üzere, değişken ve dinamik bir diğer deyişle akan bir yapıya sahip olmalarıdır. Bir yandan sisteme cihazlardan yeni veriler dahil olmakta veya bazı veriler kesintiye uğramakta, öbür yandan mevcut verilerde değişiklik meydana gelebilmektedir. Toplanan verilerin analizi bu sebeple daha karmaşık bir hal almaktadır. “Big data” yani “büyük veri” kavramı bu sebeple de özellikle son yıllarda çokça tartışılır hale gelmiştir.

İnternet ve kablosuz teknolojilerin hayatın her alanında daha çok yer almasıyla birlikte, sayılar, metinler, ifadeler, görüntüler, şekiller, grafikler, sunumlar gibi anlamlı bilgilere dönüşen verilerin devasa miktarlarda depolanması söz konusu olmaktadır. Hareket halindeki mobil uygulamalar ile birlikte, toplanan verinin hem çeşitliliği hem de hacmi inanılmaz boyutlarda artmaya devam etmektedir. Verilerin işlenmesi, sınıflandırılması, analizi, ve anlamlı bilgilere dönüştürülmesi ile kendi kendine karar veren sonuçların elde edilebilmesine de imkân vermektedir. Bu ders çalışmasında, veri analizinde yapay zekâ ve makine öğrenmesi tekniklerinin kullanımı anlatılacaktır. Başlıca yapay zekâ ve makine öğrenmesi teknikleri öğretilecektir. Kümeleme, sınıflandırma, yapay sinir ağları, metin ve web madenciliği, fikir madenciliği, duygu analizi, olasılık ve öğrenerek karar veren algoritmalar konusunda büyük verilerle yapılan çalışmalar anlatılacaktır.

Büyük Veri Analizi Büyük veri, kısaca 11V diyebileceğimiz beş kavram ile betimlenmektedir:

- 1) Volume (Hacim),
- 2) Velocity (Hız),
- 3) Variety (Çeşitlilik), 4) Verification (Doğrulama) 5) Value (Değer).
- 6) Veracity (Gerçeklik),
- 7) Volatility (Oynaklık)
- 8) Validity (Geçerlik)
- 9) Vulnerability (Hassaslık),
- 10) Variability (Değişkenlik),

11) Visualization (Görselleştirme).

“Data never sleeps” projesi kapsamında Haziran 2016 verilerine göre, yalnızca bir dakika içerisinde;

- Youtube video paylaşım sitesine kullanıcılar tarafından 400 saatlik video yüklemesinin gerçekleştirildiği,
- Twitter üzerinden 9.678 adet emoji içerikli tweet atıldığı,
- Google’ da 69.500.000 kelime tercüme edildiği, sadece Amerikalı kullanıcıların mobil cihazlarla yaklaşık 18.000 GB veri kullandıkları,
- Facebook Messenger kullanıcılarının 216.302 adet fotoğraf paylaştıkları,
- Instagram kullanıcılarının paylaşılan görüntüler için 2.430.555 adet beğeni yaptıkları
- Amazon web sitesinden 222.283 \$ satış yapıldığı tespit edilmiştir.

Büyük veriden kastedilen yalnızca hacimsel büyüklük değildir. Farklı kaynaklardan ve farklı biçimlerde toplanan verilerin anlamlı ve işlenebilir hale getirilmesi gerekmektedir. Veriler hacim ve tür yanında sürekli artan bir hızda oluşmakta ve depolanmaktadır.

Verilerin depolanma ve değişme hızının yanı sıra büyük verilerin çoğu zaman karmaşık, düzensiz olduğu ve yanlışlar içerebileceği gerçeği, bu verilerin düzenlenmesi ve ayıklanması sorununu doğurmaktadır.

Üstelik anlık alınan verilerden hemen bilginin elde edilmesi yani verinin toplandığı anda analiz edilmesi gerekmektedir. Özellikle internete bağlı cihazların kaynaklık ettiği verilerin analizinde veri madenciliği yöntemleri yanında pazarlama, algı yönetimi, izleme, web, metin ve multimedya madenciliği teknikleri kullanılmaktadır. Burada unutulmaması gereken, verilerin yalnızca dijital değil fotoğraf, resim, video, ses, metin, konum (GPS) bilgisi vs. gibi pek çok çeşitte ve her biri için çeşitli boyutlarda olduğudur. Böyle olunca da asıl önemli olan, bu kadar büyük, hızlı ve çeşitli olan veri topluluğundan anlamlı ve değerli bilgiyi elde etmek olmaktadır.