



Hafta 1. Makine öğrenmesi giriş

Makine öğrenmesi nedir?

- Makine Öğrenmesi, «**geçmiş veriler yardımıyla bugün ve gelecek için değerli bilgiler elde etmek üzere gerçekleştirilmiş bilgisayar destekli sistemler oluşturma tekniğidir**».
- Böylece bir problem, var olan önceki örneklerden çıkarım yapılarak çözülebilir.

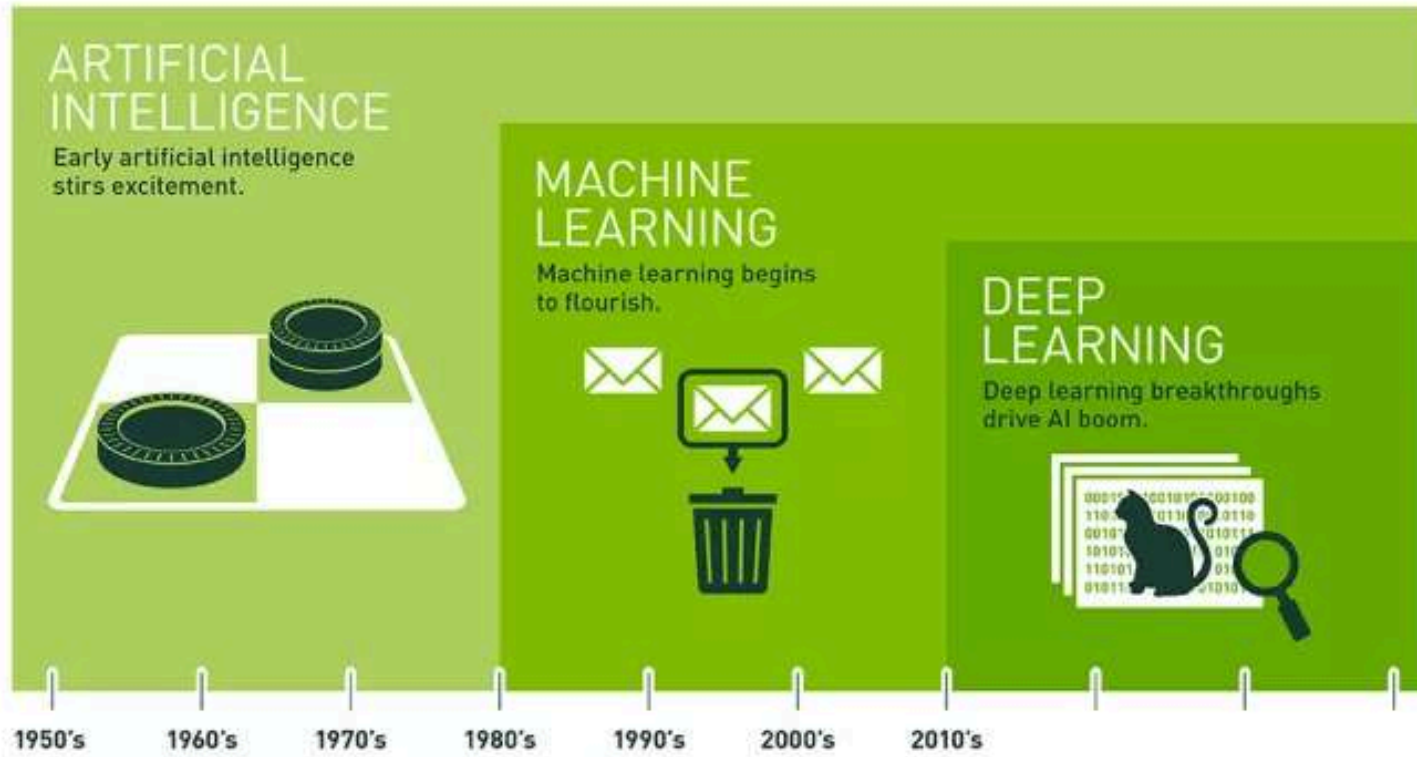
Başka bir tanım

- Yapay zekanın sayısal öğrenme ve model oluşturma çalışmalarından geliştirilmiş bir alt dalıdır.
- **Makine** öğrenmesi verilerden öğrenebilen ve yeni veriler için tahmin yapabilen algoritmaların çalışmalarını araştıran bir sistemdir.
- Makine öğrenmesi algoritmaları örnek girişlerden tahminler yapmak ve kararları gerçekleştirebilmek amacıyla bir model inşa etmeye çalışır.

Yapay zeka ile ilişkisi

- Yapay zeka, «**bir bilgisayarın veya bilgisayar kontrolündeki bir robotun çeşitli faaliyetleri zeki canlılara benzer şekilde yerine getirme kabiliyetidir**».
- Makine öğrenmesi yapay zeka sistemlerinin kendi kendine öğrenmesini ifade eder.
 - *Zeki sistemler yapay zeka ile kendini geliştiren sistemler ise makine öğrenmesi ile mümkün olacaktır.*
- Makine öğrenmesi yapay zekanın bir alt alanıdır.

Deneme → İlerleme → Patlama



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

<https://yapayzeka.ai/yapay-zeka-makine-ogrenmesi-ve-derin-ogrenme-arasindaki-fark-nedir/> Makine Öğrenmesi Ders Notları, Doç. Dr. Hidayet Takcı, Sivas Cumhuriyet Üniversitesi

Veri madenciliği ile ilişkisi

- ▶ Makine öğrenmesi algoritmaları veri madenciliği problemlerinin çözümüne yardımcı olur.
 - ▶ Örneğin sınıflandırma bir veri madenciliği problemi iken sınıflandırma işleminde kullanılan karar ağaçları bir makine öğrenmesi algoritmasıdır.
- ▶ Veri madenciliği veri ön işleme ve yorumlama dahil sürecin tamamına verilen isim iken makine öğrenmesi veri ön işlem (pre-processing) ve veri son işleme (post-processing) karışmaz.

Makine öğrenmesi nerelerde kullanılır?

- İnsan tecrübesinin olmadığı durumlarda
 - Uzay görevleri...
- Açıklanamayan becerilerimizin taklidi.
 - Yüz tanıma, konuşma tanıma...
- Duruma (zamana) bağlı çözümler gerektiğinde.
 - Bir bilgisayar ağında yönlendirme işlemi...
- Özel uygulamalar.
 - Biyometrik sistemler (parmak izi, retina)...

Örüntü tanıma

- Biz yıllar önce gördüğümüz birini yılar sonra hemen tanırız.
 - Ancak bu işlemi nasıl yaptığımızı açıklayamayız. Algoritmasını yazamayız
- Yüz görüntüsü gelişigüzel noktalardan oluşmaz. Belirli bir yapısı vardır.
 - Örneğin göz, ağız ve burunun yerleri bellidir.
- O halde;
 - Bir program bir insana ait yüz resimlerini inceleyerek o kişiye özgü örüntüyü tanımlayabilir.
 - Daha sonra verilen bir resmin bu örüntüye ne kadar benzediğini çıkarabiliriz.
- Bu işleme *örüntü tanıma* adı verilir.

Makine öğrenmesi uygulamaları

- Veri madenciliği teknikleri yardımıyla sıklıkla yapılan işlemler makine öğrenmesi uygulamaları olarak görülebilir.
- En bilinenleri şunlardır:
 - Sınıflandırma
 - Kümeleme analizi
 - Birliktelik kuralları madenciliği
 - Sıralı örüntü keşfi ve analizi
 - Anormallik tespiti

Sınıflandırma

- Yeni kayıtların daha önceden belirlenmiş sınıflara atanmasına sınıflandırma denir.
- Bir Sınıflandırma modeli, tahmin değişkenleri (predictors) yardımıyla, hedef (class) değişkeninin hesap edildiği fonksiyondur.
- Genellikle, veri seti eğitim ve test setlerine bölünür, eğitim seti ile model inşa edilir, test seti ile model doğrulama yapılır.
 - Eğitim seti (*training set*) sınıflandırma modelinin elde edileceği kayıtları içerir. Eğitim setindeki her kayıt bir tanesi sınıf (*class*) özneliği olan özneliklerden oluşur.
 - Test seti bir sınıflandırma modelinin doğruluğunu belirlemek için kullanılır.

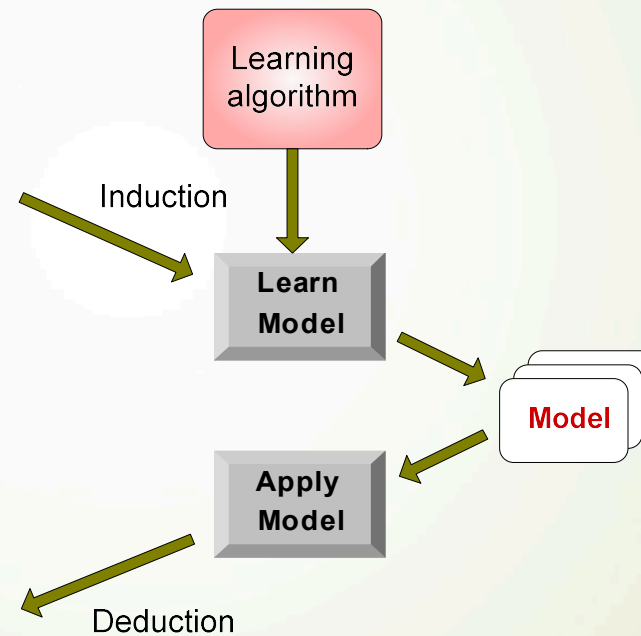
Sınıflandırma işlemi çalışma prensibi

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set



Sınıflandırma: örnek 1

- Doğrudan Pazarlama
 - **Hedef:** piyasaya yeni sürülen bir cep telefonu modelini almaya meyilli müşterileri tespit ederek reklam maliyetini düşürme.
 - **Yaklaşım:**
 - Önceden tanıtımı yapılmış benzer bir ürün için elde edilen veriler elde edilir.
 - Bu veriler içerisinde yer alan *{buy, don't buy}* bilgileri yardımıyla ürünü alacak ve almayacak müşteri türleri için model oluşturulur.
 - Böylesi bütün müşteriler için çeşitli demografik, yaşam stili, şirket bağlantıları ile ilgili bilgileri toplanır.
 - **Müşterinin işi, nerede ikamet ettiği, ayda ne kadar kazandığı, v.s.**

Sınıflandırma: örnek 2

- İstisna Tespiti (fraud detection)
 - **Hedef:** Kredi kartı işlemlerinde istisna durumların tespit edilmesi.
 - **Yaklaşım:**
 - Kredi kartı işlemleri ve kredi kartında tutulan veriler modelin nitelikleri olarak kullanılır.
 - **Müşteri ne zaman (gün, saat) alışveriş yapar, ne alır, ne sıklıkta vaktinde ödeme yapar, v.s.**
 - Geçmiş işlemler dürüst (fair) ve istisna (fraud) olarak etiketlenir. Fair ve fraud sınıf nitelikleri olsun.
 - Veriler yardımıyla model öğrenimi gerçekleştirilir.
 - Ortaya konan model aracılığıyla hesaplar üzerindeki kredi kartı işlemleri gözlemlenerek istisna tespit edilir.

Sınıflandırma: örnek 3

- Müşteri Memnuniyeti/Memnuniyetsizliği :
 - **Hedef:** Bir müşterinin bir rakibe kaptırılma olasılığı olup olmadığı tahmin edilmeye çalışılır.
 - **Yaklaşım:**
 - Nitelikleri bulmak için eski ve şu andaki müşterilerin her birinin detaylı işlem kayıtları kullanılır.
 - Müşteri çağrıları ne sıklıkta, en çok günün hangi saatinde çağrı yapılır, müşterinin finansal durumu ve medeni durumu nedir, vs.
 - Müşteriler **sadık** veya **sadık değil** diye etiketlenir.
 - Sadakat için bir model bulunur.

Sınıflandırma için kullanılabilecek makine öğrenmesi algoritmaları

- Karar ağaçları (ID3, Rnd Tree, J48)
- Kural tabanlı sınıflayıcılar (C4.5, C-RT, Decision Table)
- Örnek tabanlı sınıflayıcılar (k-NN, KStar, LWL)
- Bayes sınıflayıcı (Naive bayes, Naive bayes continuous, Bayes Net)
- Yapay sinir ağları (Multilayer perceptron, Perceptron, LVQ)
- Destek vektör makineleri (CVM, C-SVC, SMO)

Benzerlik/uzaklık tabanlı sınıflayıcılar

- Benzerlik/uzaklık sıklıkla kümelemeyle ilgili bir kavram olsa da sınıflayıcılarla da kullanılır.
- Benzerliğin kullanıldığı en önemli örneklerden birisi CENTROID sınıflayıcıdır. Centroid sınıflayıcılar genellikle doküman sınıflandırmada kullanılır.
- Her bir sınıf bir centroid vektörü ile sunulurken, her bir doküman bir doküman vektörü ile sunulur.
- Sınıflandırma işlemi doküman vektörü ile sınıf vektörleri arasındaki benzerlik/uzaklıklardan elde edilir.
- Benzerlik ve uzaklık için verinin türüne göre birçok yöntem vardır. En bilinenleri; Euclidian, Manhattan, Jaccard, Cosine v.s. dir.

Sınıflandırma ve regresyon

- Sınıflandırma ve regresyon aynı görevi yerine getirdiği için sıklıkla regresyon da bir sınıflayıcı olarak ele alınır.
- Aralarındaki en önemli fark hedef değişkenle ilgilidir;
- Şöyle ki;
 - Eğer hedef değişkenin türü NOMINAL (categorical) ise SINIFLANDIRMA
 - Eğer hedef değişken türü SÜREKLİ (continuous) ise REGRESYON
- Dolayısıyla regresyon, sonuçta sürekli değer üretir. Sadece lojistik regresyon nominal değer üretir.

Kümeleme

- Elimizde, verileri sunan noktalar ve bu noktalar arasındaki benzerliği ölçen bir benzerlik ölçümü varken, kümelemenin amacı; aşağıdaki özellikleri sağlayan kümeleri bulmaktır:
 - Birbirine daha çok benzeyen noktalar aynı kümede,
 - Birbirine benzemeyen noktalar ise farklı kümelerdedir
- Benzerlik Ölçümleri:
 - Eğer öznitelikler sürekli değerler ise o zaman Euclidean Distance.
 - Diğerlerinde probleme uygun ölçümler kullanılır.

Kümeleme: örnek 1

► Pazar Bölümlemesi:

► **Hedef:** Her biri ayrı bir pazarlama stratejisi için uygun olabilecek farklı müşteri alt gruplarını bulmak.

► Yaklaşım:

- Müşterilerin yerleşim bölgelerine ve yaşam stillerine dayalı farklı özellikleri toplanır.
- Benzer müşterilerin kümeleri bulunur.
- Aynı kümedeki müşteri örüntüleri gözlemlenip diğer kümelerle karşılaştırma yapılarak kümeleme kalitesi ölçülür.

Kümeleme: örnek 2

► Doküman Kümeleme:

- **Hedef:** İçerisinde geçen önemli terimlere dayalı olarak benzer dokümanların gruplarını bulmak.
- **Yaklaşım:**
 - Her bir dokümanda kullanılan terimlerin frekansları belirlenir.
 - Terimlerin frekanslarına dayalı olarak bir benzerlik bulunur.
 - Bu bilgi kümeleme için kullanılır.

Kümeleme algoritmaları

- Merkez tabanlı kümeleme (k-means, k-medoids)
- Hiyerarşik kümeleme (Hierarchical clusterer)
- YSA ile kümeleme (Kohonen-SOM)

Birliktelik kuralları

- Bir işlem kaydında bir elemanın meydana gelme olasılığını, diğer elemanların meydana gelme olasılıklarından tahmin etmek için kurallar bulunmasıdır.

<i>TID</i>	<i>Items</i>
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

$\{\text{Diaper}\} \rightarrow \{\text{Beer}\},$
 $\{\text{Milk, Bread}\} \rightarrow \{\text{Eggs, Coke}\},$
 $\{\text{Beer, Bread}\} \rightarrow \{\text{Milk}\},$

Birliktelik kuralları: örnek 1

- Supermarketin kendi kendine yönetimi
 - **Hedef:** Birçok müşteri tarafından birlikte satın alınan elemanların belirlenmesi.
 - **Yaklaşım:** Elemanlar arasındaki bağlantıları bulmak için barkod tarayıcılar ile toplanan satış noktası verileri işlenir.
 - Bir klasik kural --
 - Eğer bir müşteri bez ve süt satın almışsa, büyük ihtimalle bira da alır.
 - Böylece, bez, süt ve bira yakın raflara konmalıdır

Birliktelik kuralları: örnek 2

► **Envanter Yönetimi:**

- **Hedef:** Bir müşteri araç bakım şirketi onun müşterilerine ait ürünlerin olası onarımlarını tahmin ederek ona uygun şekilde araç ekipmanlarını elde tutmak istemektedir.
- **Yaklaşım:** farklı müşteri yerleşimlerinde önceki onarımlarda ihtiyaç duyulan parça ve araçlar hakkındaki veri işlenir ve örüntüler arasındaki tekrarlar keşfedilir.

Birliktelik kuralları için makine öğrenmesi algoritmaları

- Apriori
- Apriori MT
- Frequent itemsets


Sıralı örüntü keşfi

- Zamana bağlı olarak birlikte gelişen olayların tespit edilmesidir. Meydana gelen bir veya birden fazla olay ondan sonra gelecek yeni bir olayın habercisidir.
- Kurallar zaman bilgisi ile sunulmaktadır.
- Birliktelik kurallarının özel bir uygulamasıdır.

(A B)	(C)	(D E)
-------	-----	-------

Sıralı örüntü keşfi: örnekler

- Satış noktası için işlem dizileri,
 - Bilgisayar Kitapçısı:
(Visual C'ye Giriş) (Yeni Başlayanlar için C++) --> (Sistem Programlama)
 - Spor giyim dükkanı:
(Ayakkabılar) (Raket, Raket Topu) --> (Spor Kıyafeti)
- Telekomünikasyon alarm loglarında,
 - (Ters Çevirme Problemi, Aşırı Akım)
(Doğrultucu Alarmı) --> (Yangın Alarmı)



Hafta 2. Makine öğrenmesi algoritmaları