Machine Learning Giriş

BTK Akademi

Büyük Veri Dönemi

- İnsanlığın ürettiği veri miktarı eksponansiyel olarak artmakta.
- Üretilen verinin karmaşıklığı da veri miktarıyla birlikte artmakta.
- Mevcut veri miktarını insan eliyle incelemek maliyetli, zaman alıcı ve istenen performansı vermemekte.
- Mevcut veri miktarını insan eliyle incelemek maliyetli, zaman alıcı ve istenen performansı vermemekte.
- Yeni teknolojiyle birlikte büyük miktarda verileri işleyebilecek işlemciler ucuza üretilebilmekte.
- Makine öğrenimi, bu yeni dönemin olmazsa olmaz bir parçası olmuş durumda.

Büyük Veri Kullanım Alanları

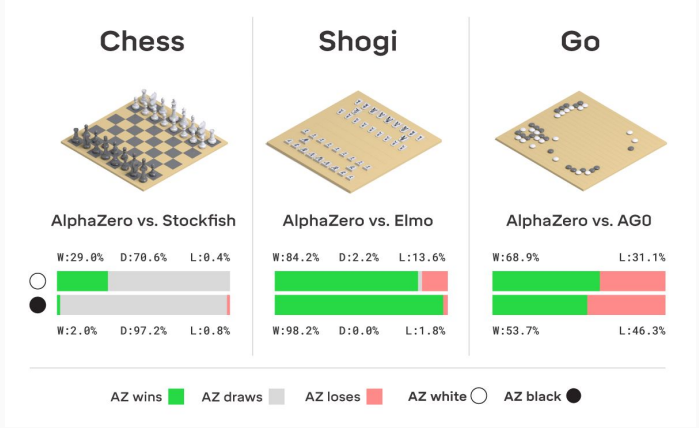
Pek çok sektörde düzenli olmayan ve insan eliyle işlenemeyen büyük miktarda veri bulunmakta.

- Ticari Faaliyetler
  - Speech recognition
  - Reklamcılık, kişiye özel pazarlama, kişiye özel sağlık hizmetleri
- Güvenlik Faaliyetleri
  - Spam mail tespiti, dolandırıcılık faaliyetleri tespiti
  - Görüntü işleme, gözetleme, takip faaliyetleri
- Bilimsel Faaliyetler
  - Ekonomik faaliyetler, piyasa tahmini
  - Biyoloji, astronomi, jeoloji, neuroscience, fizik, vb.

Büyük Veri Kullanım Alanları

Speech recognition: http://y2u.be/JvbHu_bVa_g
Görüntü işleme: http://y2u.be/50NPqEla0CQ
İniş yapan roketler: http://y2u.be/gsliniJMr3E

STM TOGAN



CNBC MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV WATCHLIST CRAMER PRO

80% of the stock market is now on autopilot

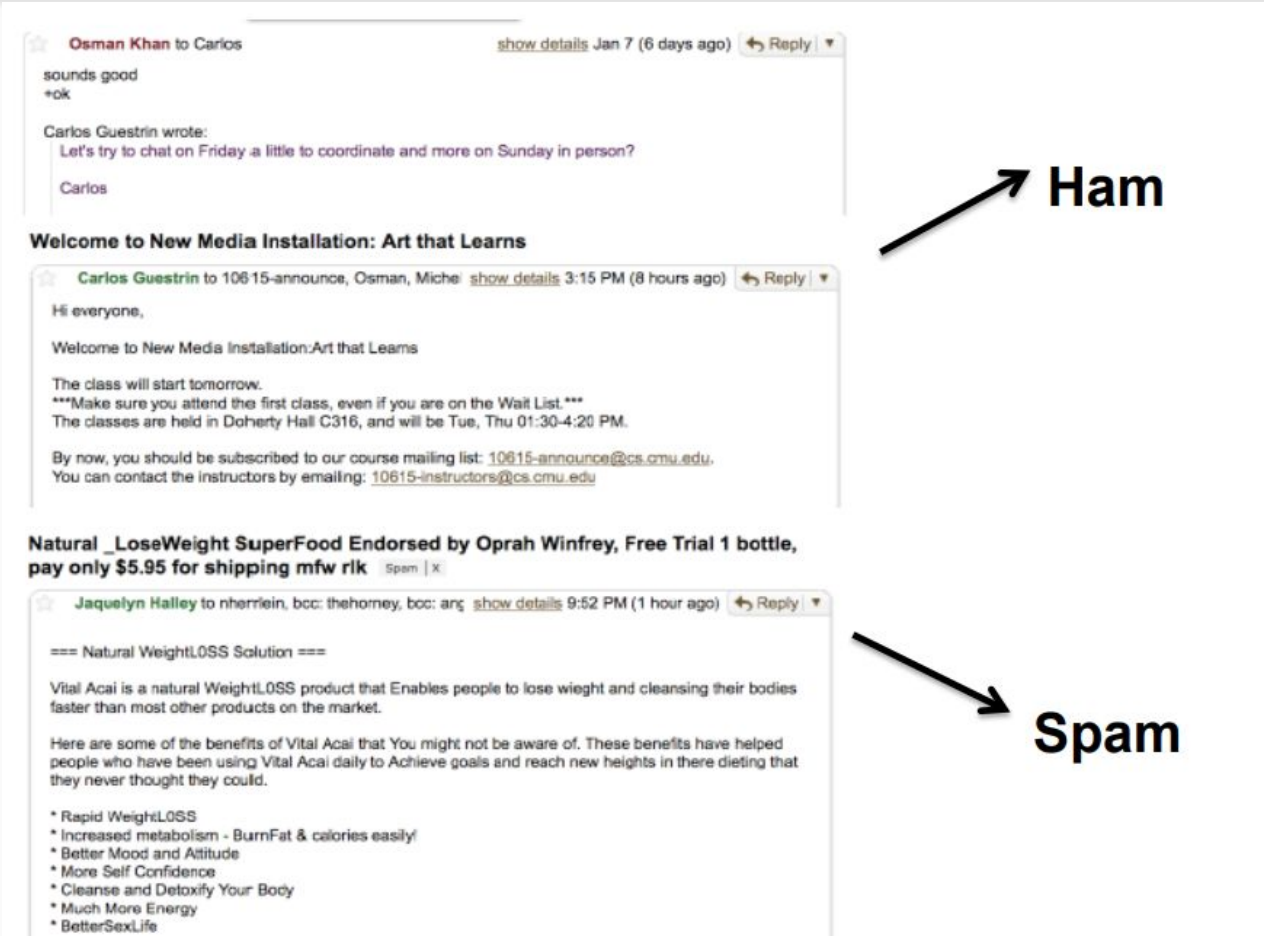
PUBLISHED SAT, JUN 29 2019•8:30 AM EDT | UPDATED SAT, JUN 29 2019•8:31 AM EDT

Yun Li
@YUNLI626

SHARE

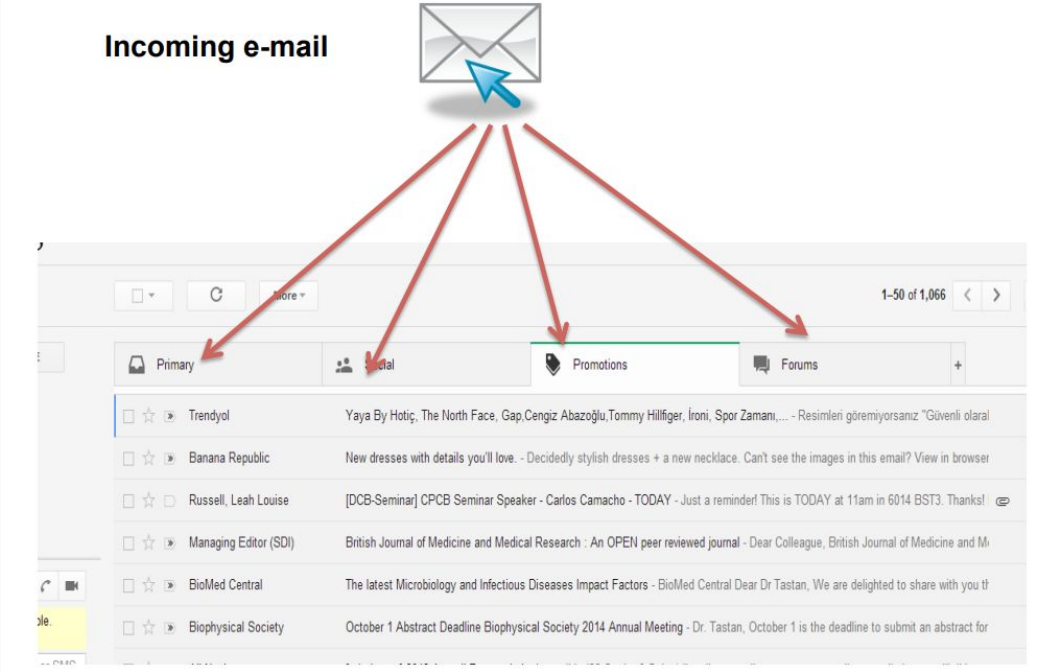
Büyük Veri Kullanım Alanları

Spam Mail Tespiti



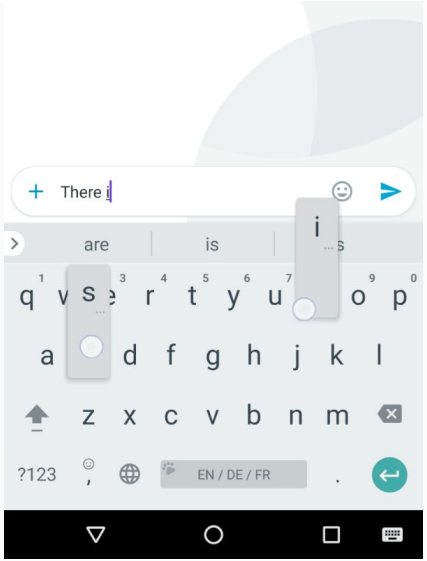
Büyük Veri Kullanım Alanları

Metin Sınıflandırma



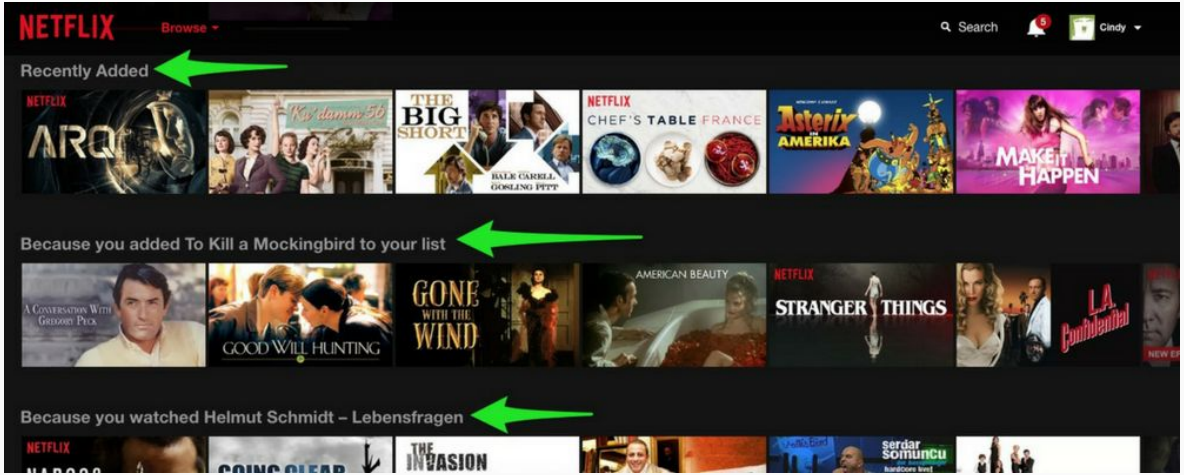
Varlık İsmi Tanıma

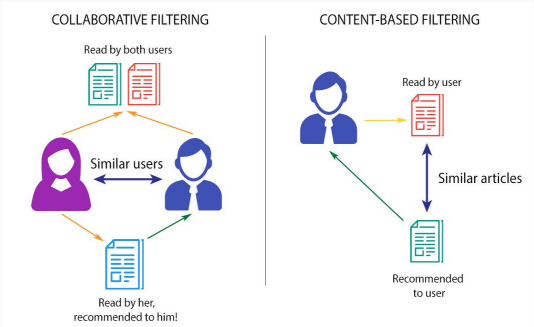
Autocomplete

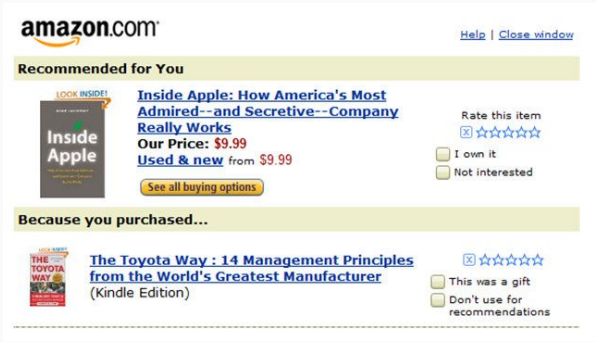


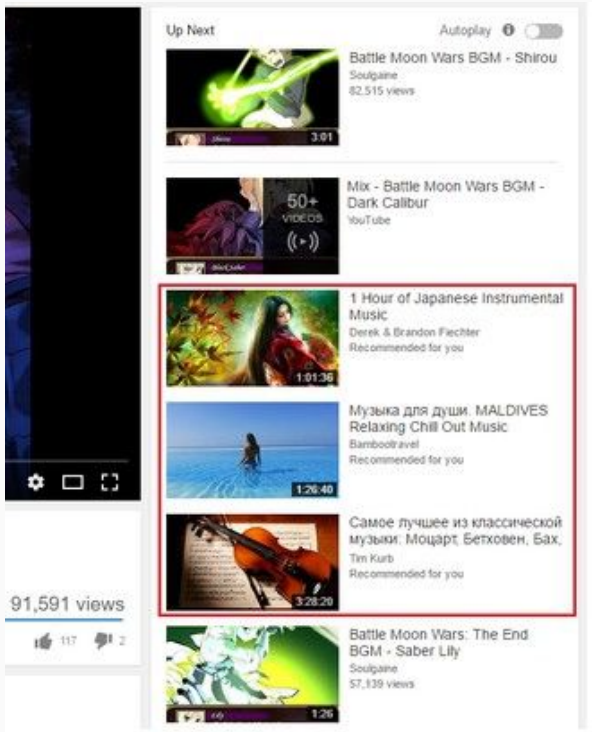
Büyük Veri Kullanım Alanları

Öneri Sistemleri



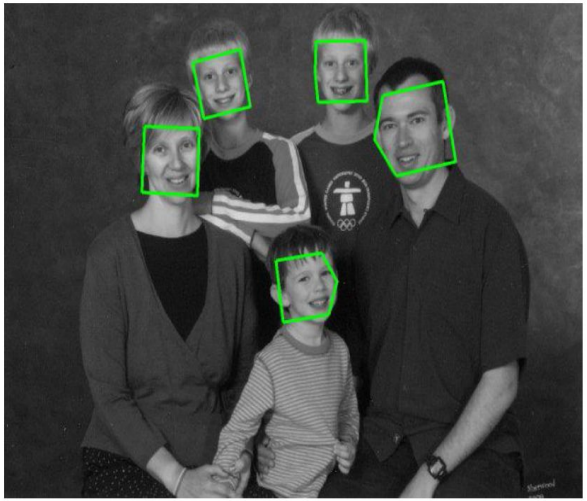


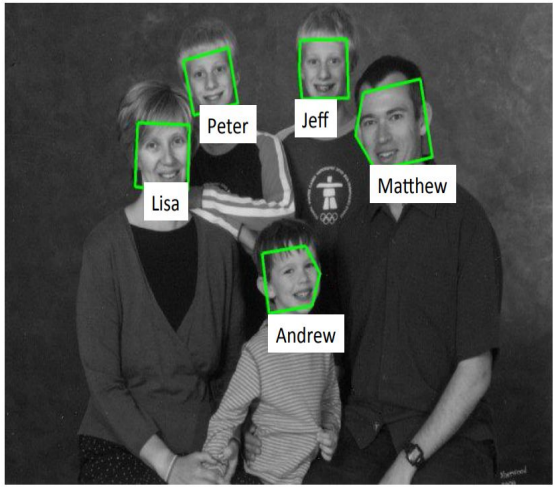




Büyük Veri Kullanım Alanları

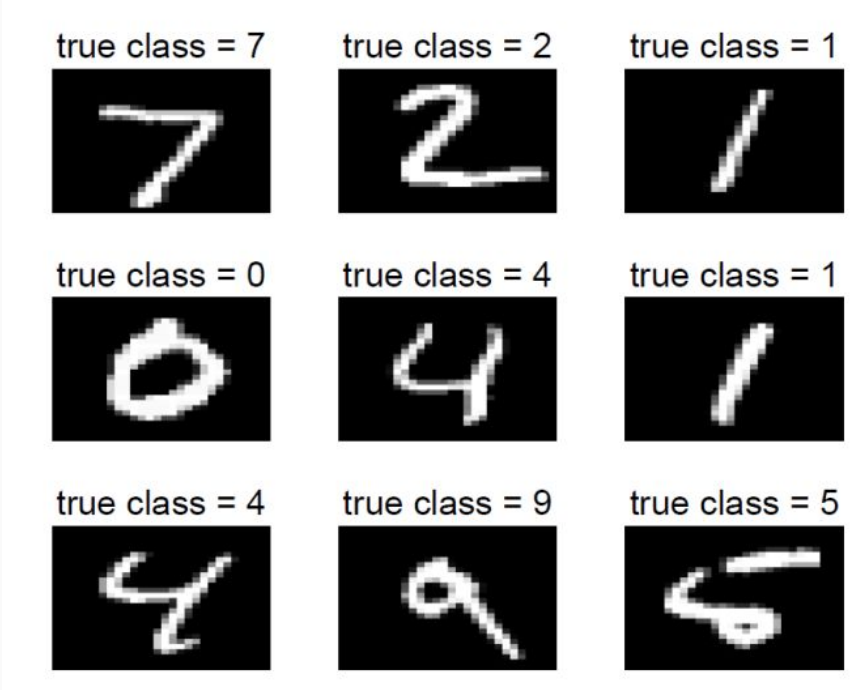
Yüz Tespiti





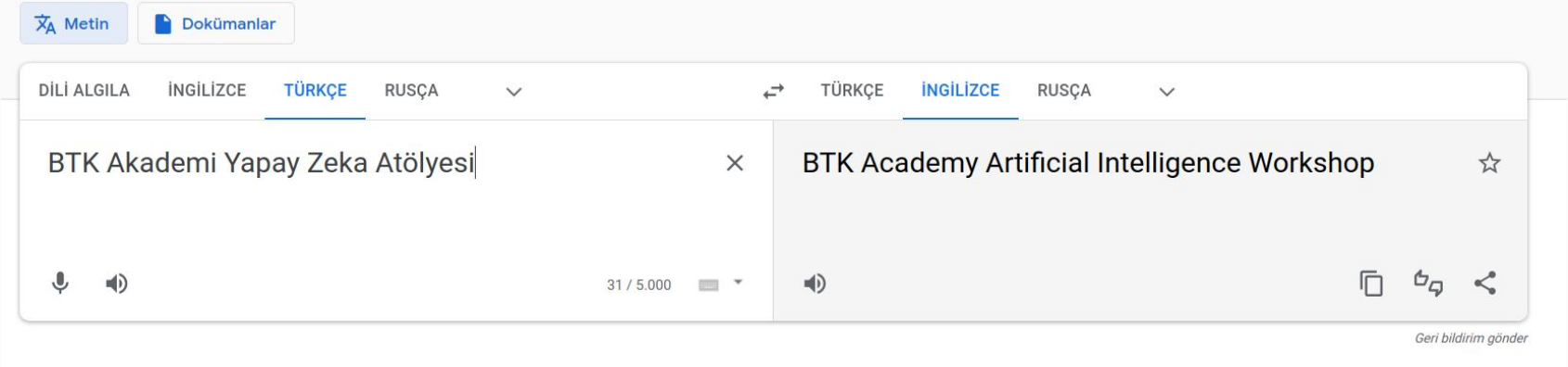
Büyük Veri Kullanım Alanları

Resim Sınıflandırma



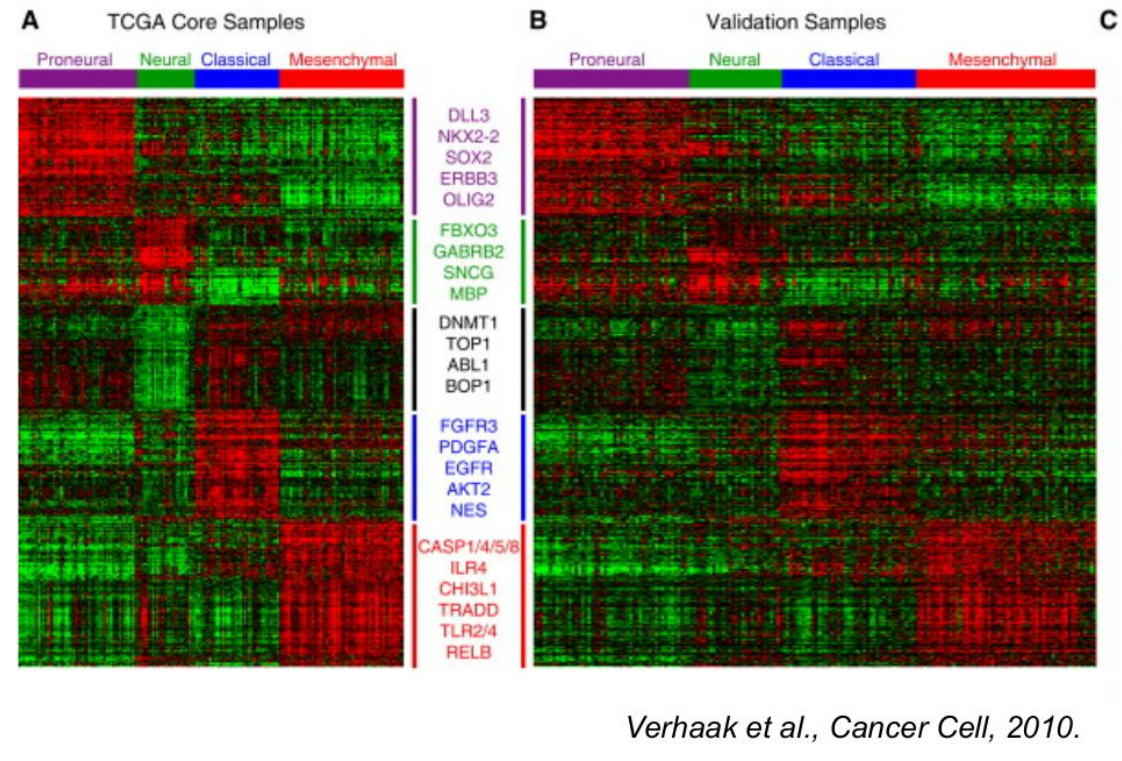
Büyük Veri Kullanım Alanları

Makine Çevirisi



Büyük Veri Kullanım Alanları

Kanserli Hücre Tespiti



Verhaak et al., Cancer Cell, 2010.

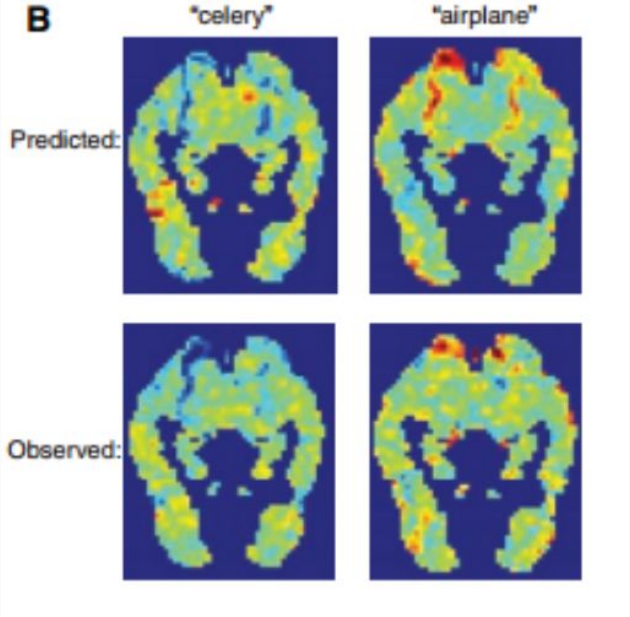
Büyük Veri Kullanım Alanları

Görselden Metne Çevirme

Automatically captioned: “Two pizzas sitting on top of a stove top oven”

Büyük Veri Kullanım Alanları

FMRI Görsellerinin Analizi / Hastalık Tespiti



Büyük Veri Kullanım Alanları

Sağlık

Büyük Veri Kullanım Alanları

Görüntü Yaratma

Büyük Veri Kullanım Alanları

Deepfake

https://www.youtube.com/watch?v=F4G6GNFz0O8&ab_channel=DiepNep

Star Wars: Rogue One, Princess Leia (Carrie Fisher)

The Mandalorian: Book of Boba Fett, Luke Skywalker (Mark Hamill)

Makine Öğrenimi

Makine öğrenimi ne zaman kullanışlıdır?

- İnsan bilgisi ve uzmanlığı yeterli değil.
  - Marsa yolculuk
  - İnsanüstü satranç becerisi

- İnsan bilgisi ve uzmanlığı var, fakat modellenebilir değil.
  - Speech-to-text
  - speech recognition
  - image recognition

- Belirli bir alanda insan bilgisi ve uzmanlığı var, fakat farklı bir alana aktarılmak isteniyor.
  - Kişiselleştirilmiş araçlar

Makine Öğrenimi

Makineler nasıl “öğrenir”?

- İnsanlar yeteneklerini geliştirmeyi zaman içerisinde pratikle öğrenirler.

- Makine öğrenmesinin amacı **tecrübe** ile gelişebilen algoritmalar kurmaktır.
  - Makine öğrenmesi algoritmaları için tecrübe **verilerden** gelir.

- **Veriler** nereden gelir?
  - İnsanlar tarafından yapılan veri etiketlemeleri
  - Otonom veri toplama/üretme

Makine Öğrenimi

Makine öğrenimi amaçları nelerdir?

- Algoritmalar:
  - Büyük çapta problemlere hızlı çalışan, yüksek başarımlı, genelleştirilmiş çözümler sunmak
  - Daha önce görmediği örneklerde de başarılı tahminler yapmak
  - Farklı alanlarda ve problemlerde kullanılabilir olmak

Tanımlar

Örnek:
- Veri içerisindeki tek bir obje veya sayır

Feature (Öznitelik) (X ile gösterilir):
- Bir örneğe ait olan, genellikle numerik formda bulunan özellik. Örneğe ait bilgi. Örn. bir kişiye ait boy, yaş, kilo, cinsiyet

Label (Etiket) (Y):
- Öznitelikler aracılığıyla tahmin edilmek veya bulunmak istenen bilgi.
- Sınıflandırma problemlerinde örneğin ait olduğu kategori bilgisi
- Regresyon problemlerinde numerik bir değer

Tanımlar

Eğitim veti seti:

- Model eğitimi amacıyla kullanılan veri seti

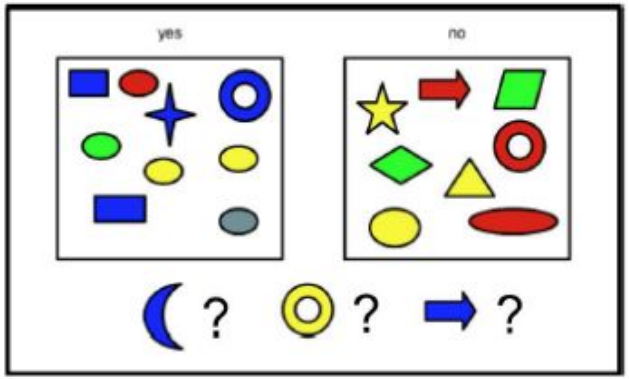
Test veri seti:

- Model başarımını ölçmek amacıyla kullanılan veri seti

Tanımlar

Feature Extraction (Veri Çıkarma):

- Ham haldeki veri içerisinden çıkarılan anlamlı bilgiler, veriyi işlenebilir hale getirme, veri içerisinden öznitelikler yaratma işlemi



(a)

D features (attributes)

N cases

Color	Shape	Size (cm)	Label
Blue	Square	10	1
Red	Ellipse	2.4	1
Red	Ellipse	20.7	0

(b)

Tanımlar

Feature Extraction (Veri Çıkarma):

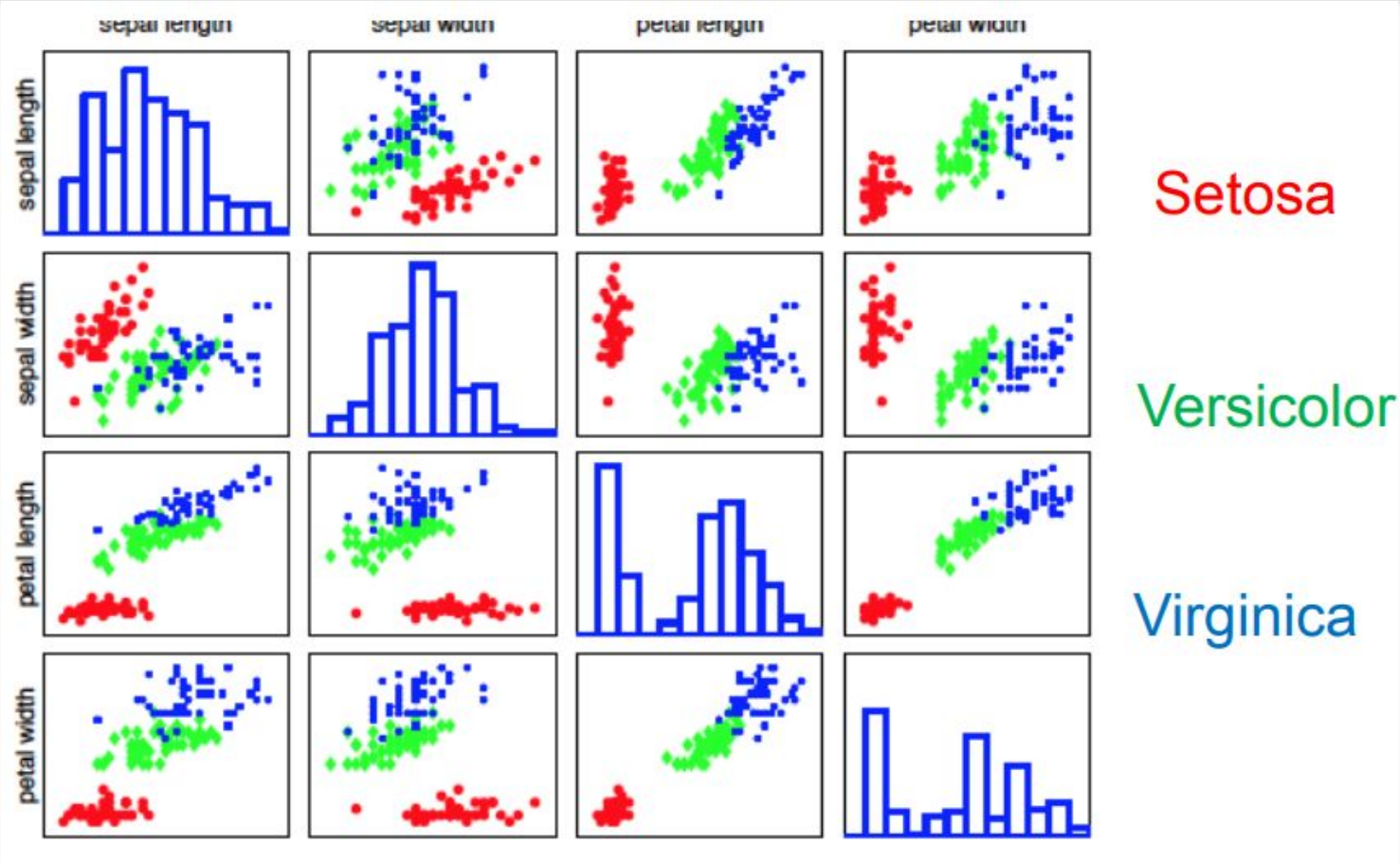
Örnek: Çiçek sınıflandırma

Setosa

Versicolor

Virginica

Tanımlar


sepal length
sepal width
petal length
petal width
sepal length
sepal width
petal length
petal width
Setosa
Versicolor
Virginica

Tanımlar

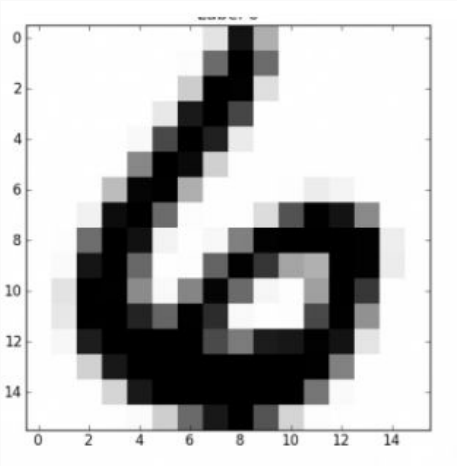
Feature Extraction (Veri Çıkarma):

Örnek: El Yazısı Tanıma

Tanımlar

Feature Extraction (Veri Çıkarma):

Örnek: El Yazısı Tanıma



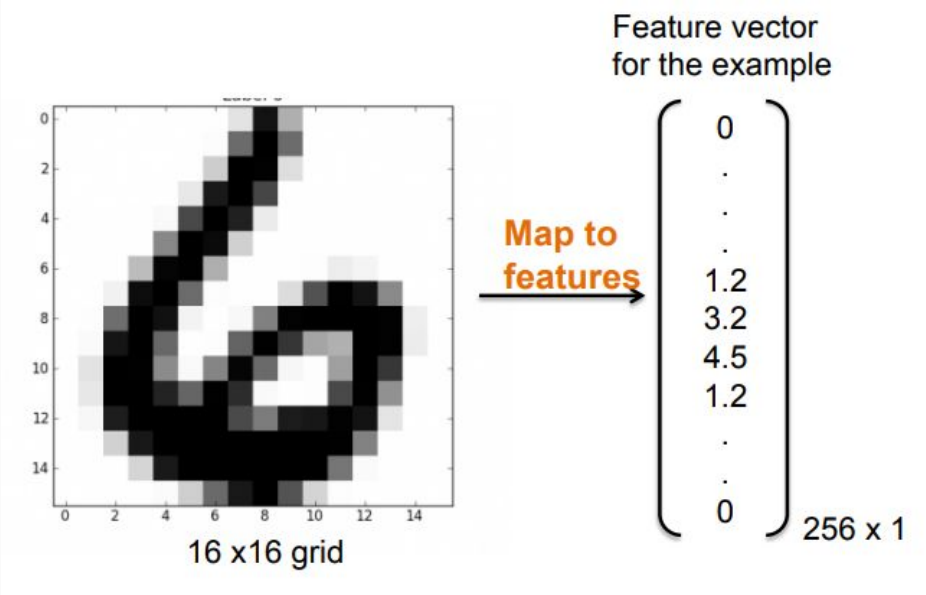
$$Y \in \{0, 1, \ldots, 9\}$$

Bu örneğin etiketi: 6

Tanımlar

Feature Extraction (Veri Çıkarma):

Örnek: El Yazısı Tanıma



Tanımlar

Training set

$$D_{train} = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$$

+ their labels

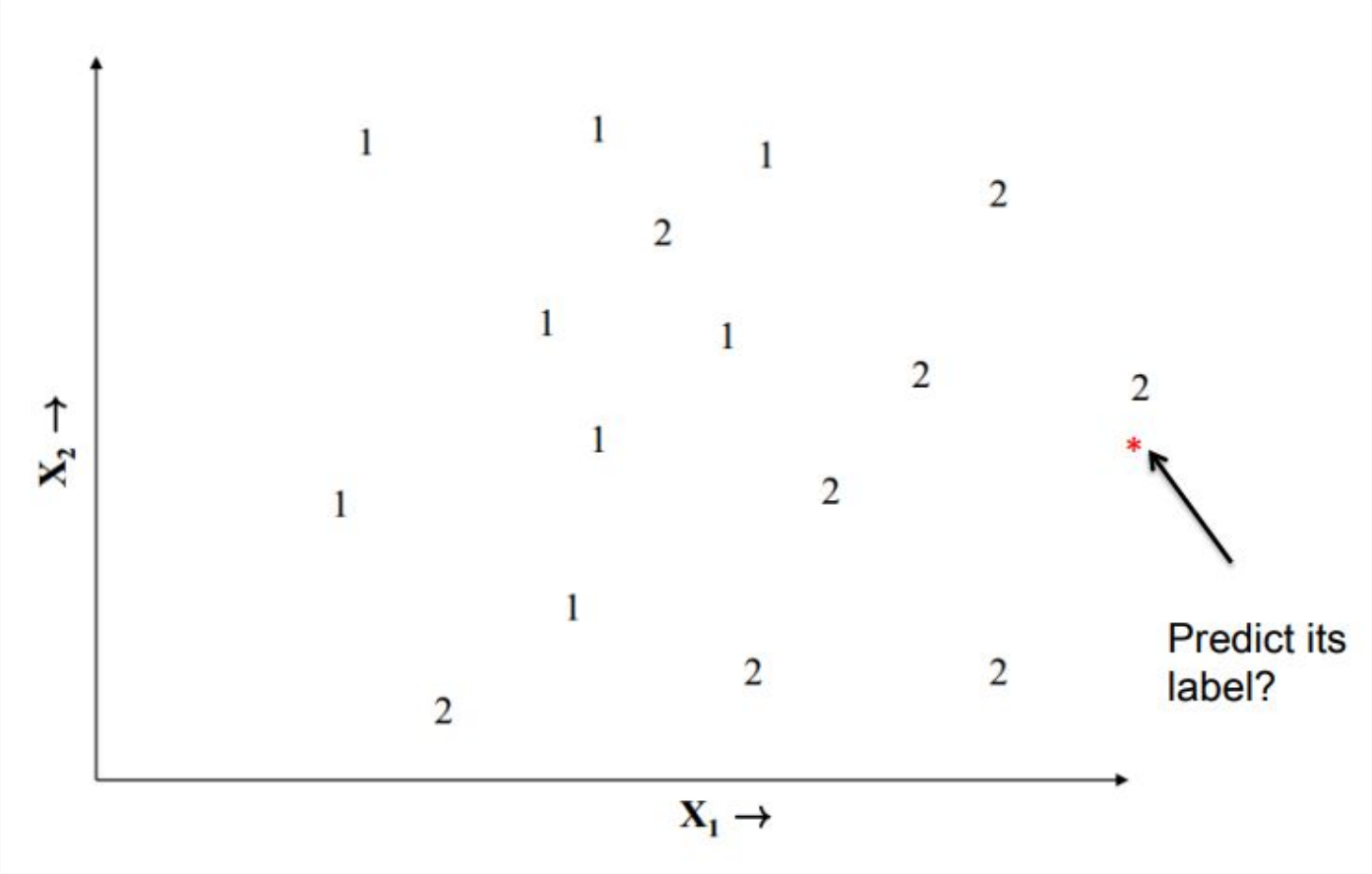
Test set

$$D_{test} = \{(X_1, Y_1), \ldots, (X_m, Y_m)\}$$

+ their labels

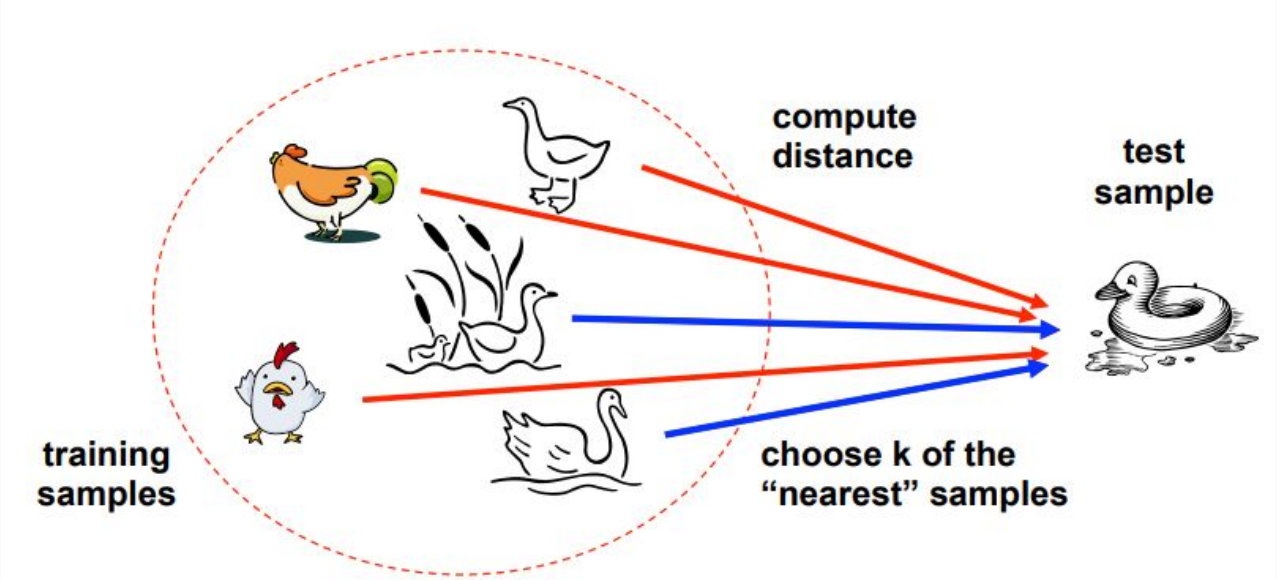
Should not be overlapping

Bu Örneği Nasıl Sınıflandırırsınız?

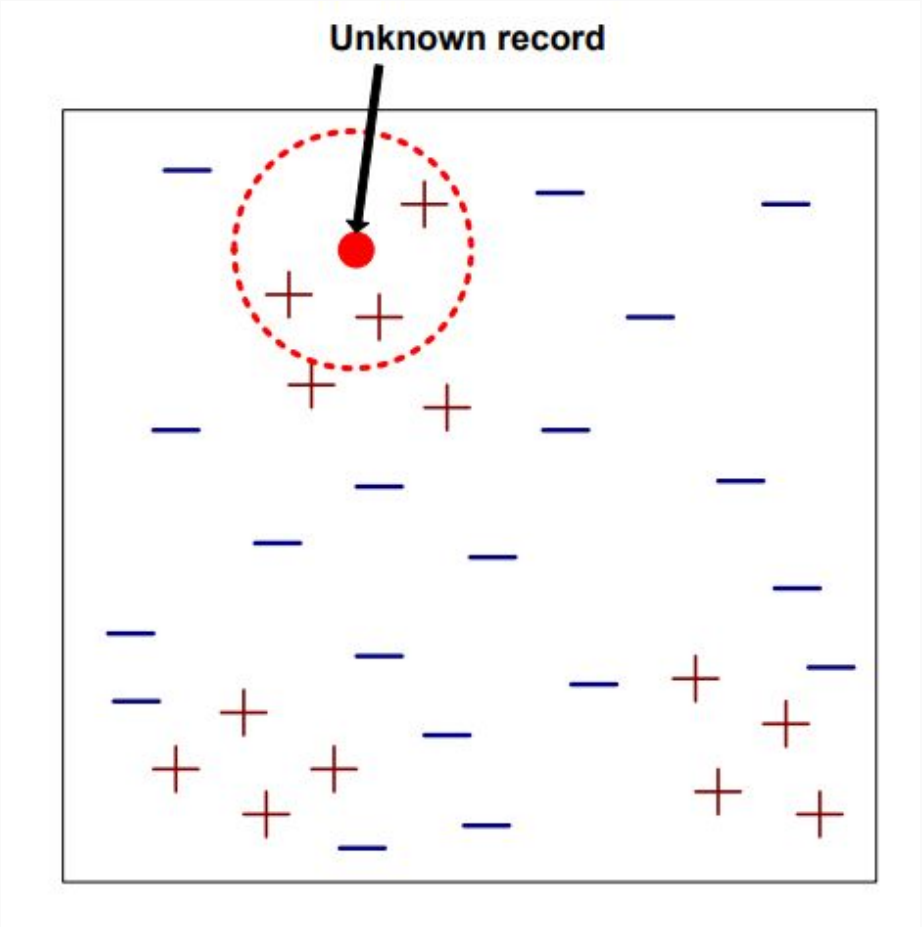


Nearest Neighbor Classifiers

Temel fikir: Ördek gibi yürüyorsa, ördek gibi ses çıkarıyorsa, ördek gibi yüzüyorsa, o zaman bir ördektir.

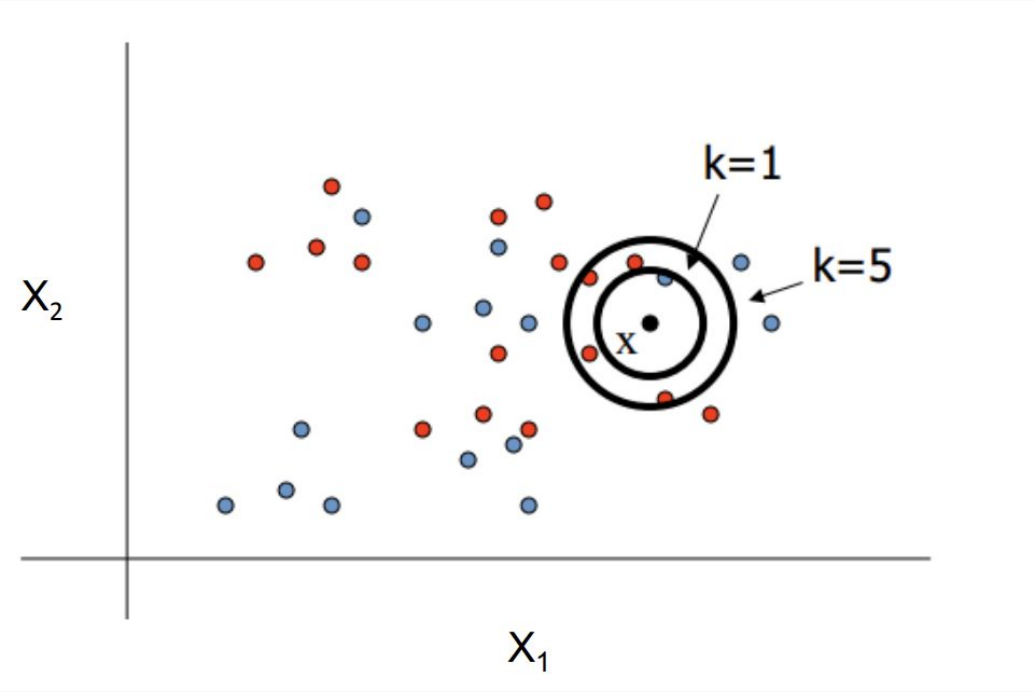


Nearest Neighbor Classifiers

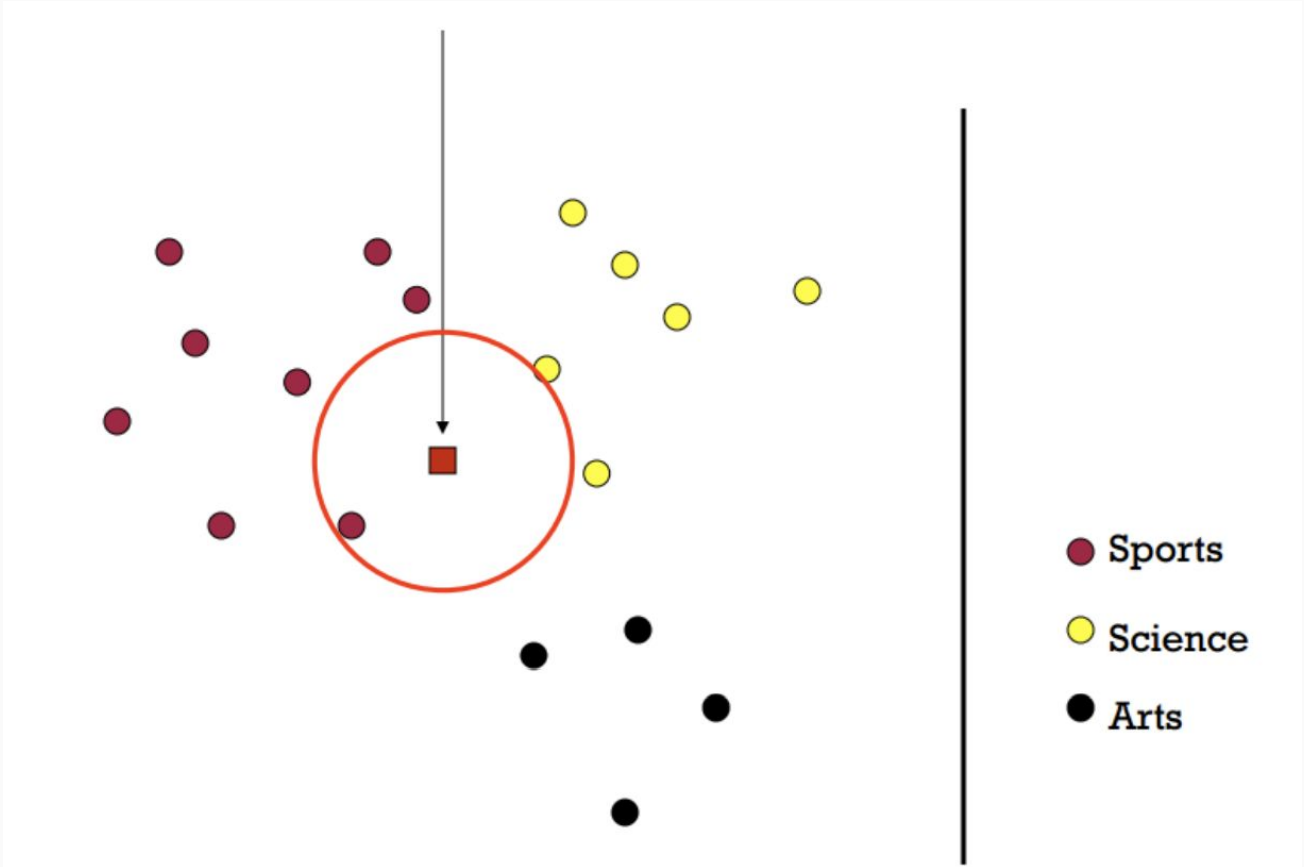


KNN Sınıflandırıcı

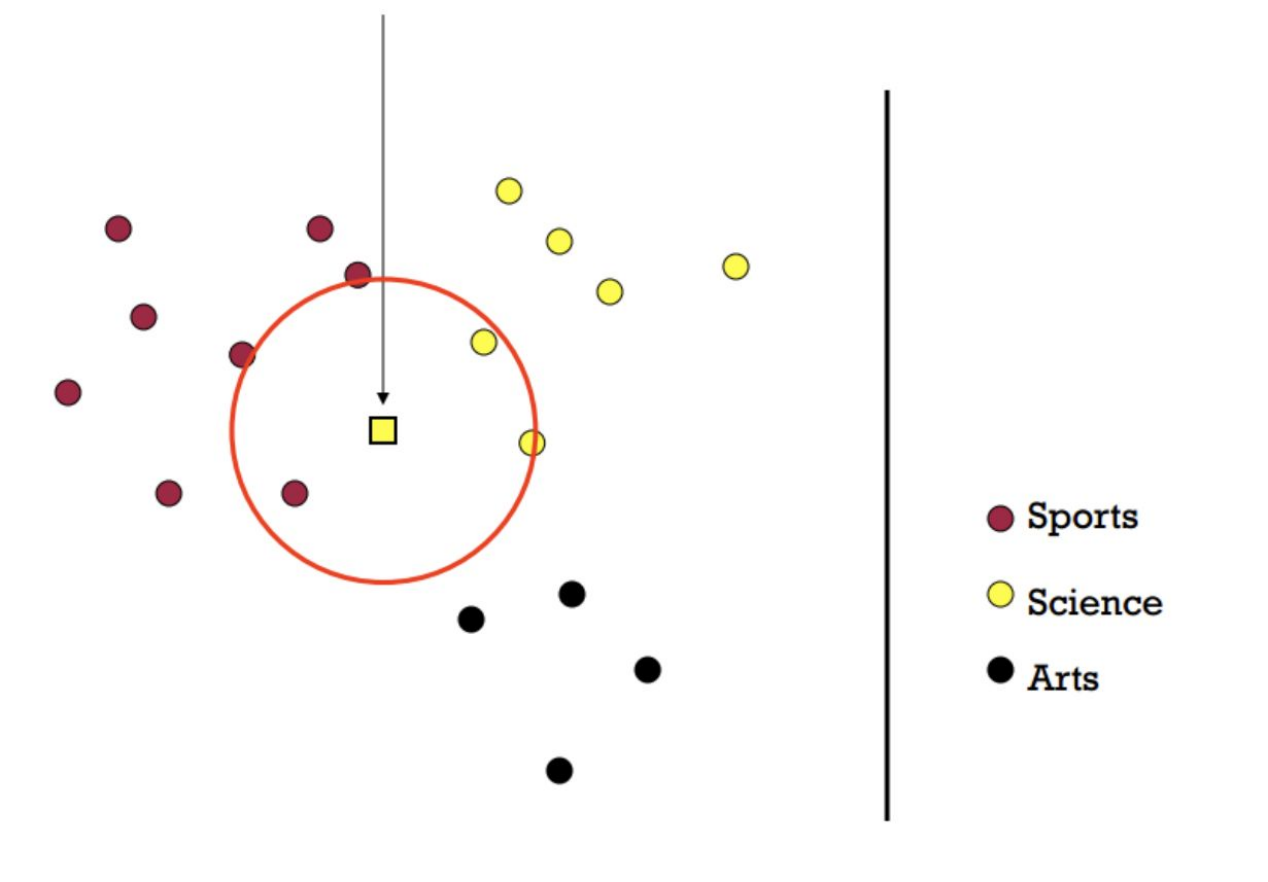
Gelen bir örneği sınıflandırmak için örneğin vektör uzayındaki konumunu analiz et ve en yakınındaki örneklere göre bir sınıf ata.



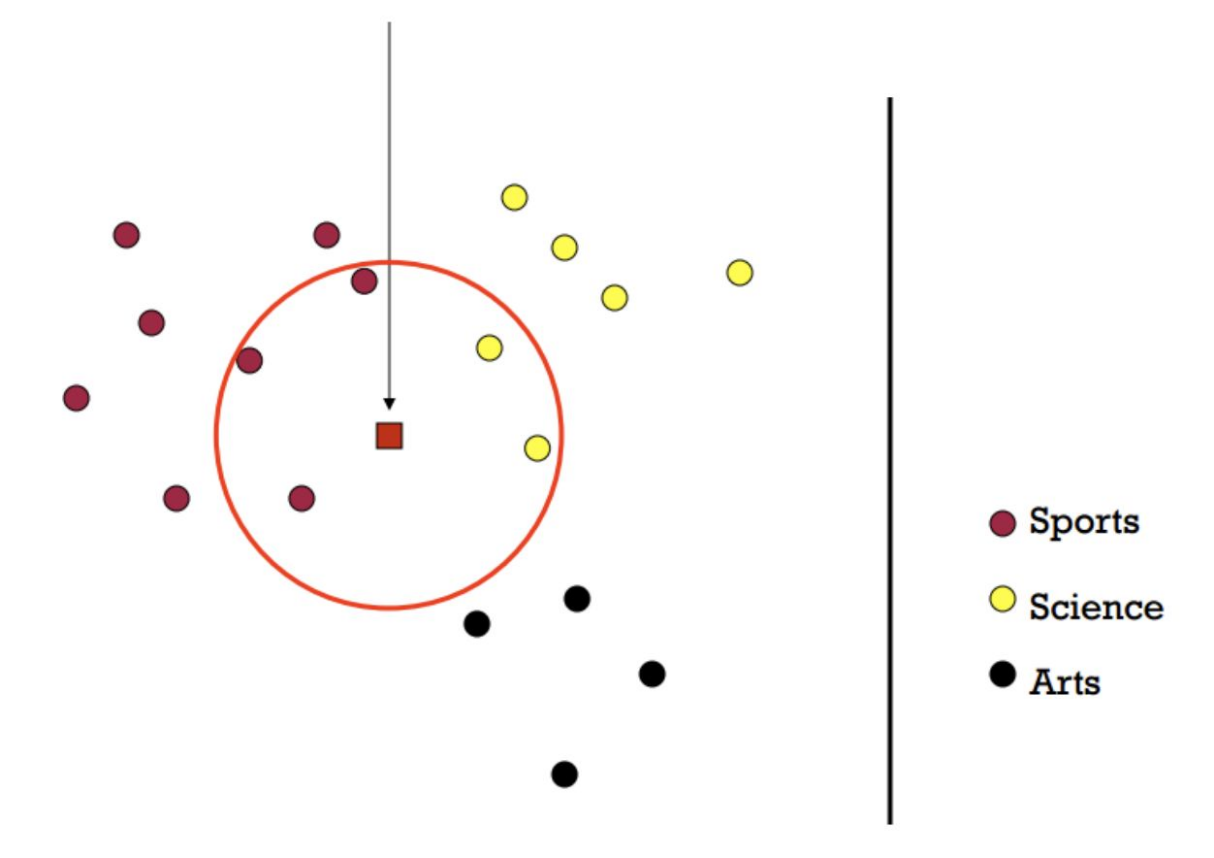
1-NN (K=1)



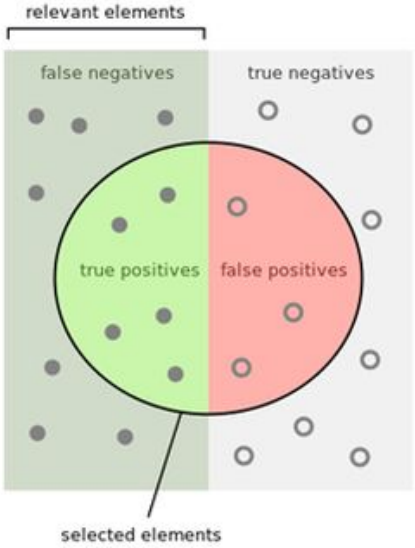
3-NN (K=3)



5-NN (K=5)



Performans Metrikleri



How many selected items are relevant?

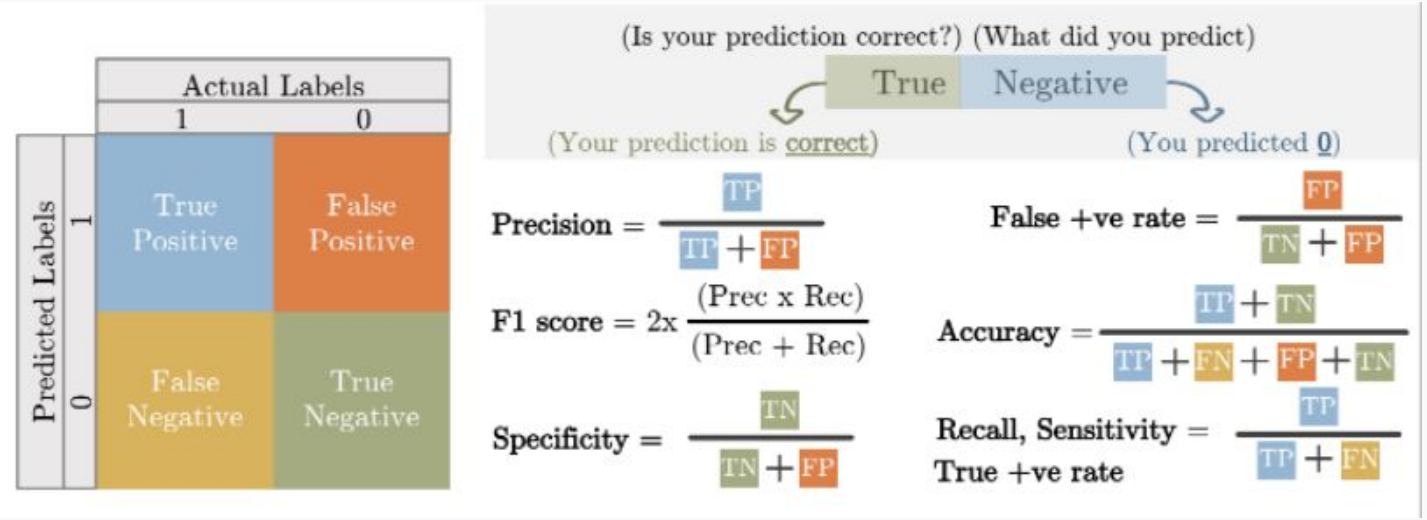
$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

Predicted Labels \ Actual Labels	1	0
1	True Positive	False Positive
0	False Negative	True Negative

Performans Metrikleri



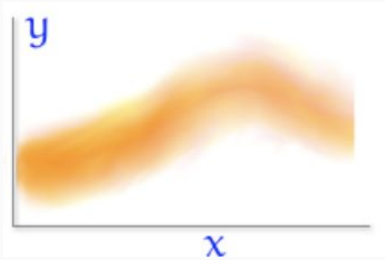
Makine Öğrenimi - Problemlerin Modellenmesi

Fonksiyonel Temsil

Problemi modellerken amaç öznitelikler (X) ile etiketler (Y) arasında bir fonksiyon öğrenebilmektir.

Örneğin:
X = [yaş, kilo, kan şekeri oranı, …], y= {diyabet hastası, sağlıklı}

Modelleme yapılırken X ile Y arasında bir ilişki olduğu varsayımı yapılmaktadır.
Örnek hipotez: yaşı yüksek, kilosu fazla ve kan şekeri oranı yüksek kişilerin diyabet hastası olma olasılığı daha yüksek olabilir.
Bu ilişki matematiksel olarak P(X, Y) olarak ifade edilebilir.



Farklı Öğrenme Yöntemleri

Etiketin var olup olmamasına göre öğrenme biçimleri:

Supervised Learning: $\{(X_1, Y_1), (X_2, Y_3), \ldots, (X_n, Y_n)\}$

Unsupervised Learning: $\{(X_1, X_2, \ldots, X_n\}$

Semi-supervised Learning: ikisinin karışımı.

Farklı Öğrenme Yöntemleri

Etiket Tipleri:

- Binary Classification $Y \in \{0, 1\}$
- Multi class Classification $Y \in \{0, 1, ...K\}$
- Regression $Y \in \Re$
- Structure Prediction Y kompleks bir yapıya sahip (graph, multi dim array vb.)

Regresyon problemleri

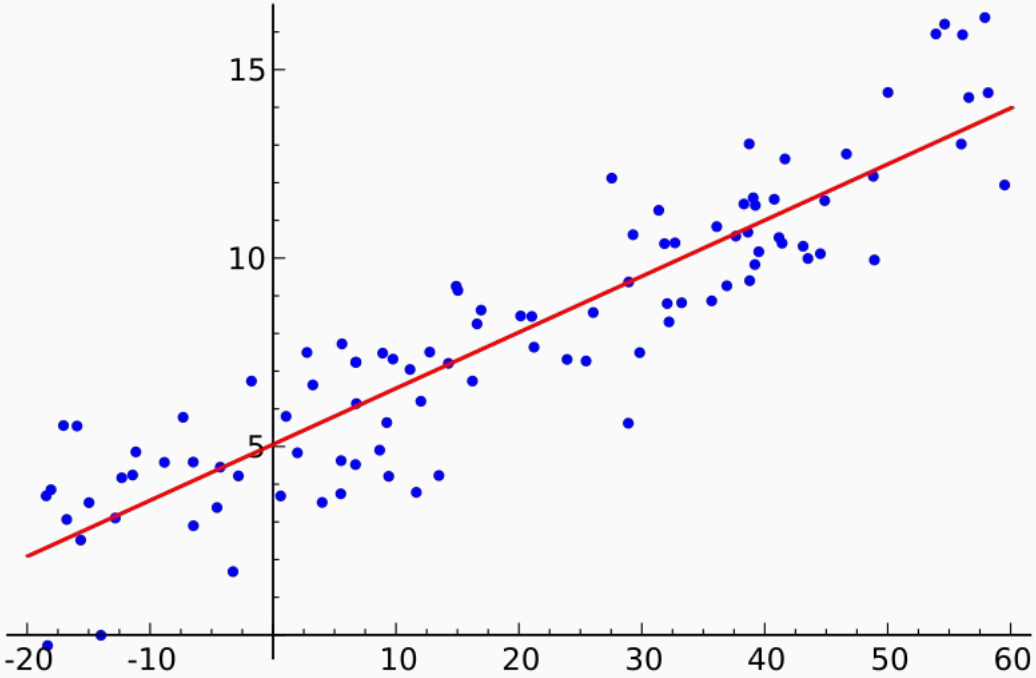
- Geçmiş verilerden yola çıkarak önümüzdeki haftaki bitcoin fiyatını tahmin etmek.
- Youtube geçmişinden yola çıkarak kişinin yaşını tahmin etmek.
- Dinlediği müziklerden yola çıkarak kişinin ne kadar mutlu olduğunu tahmin etmek.
- Bir kişinin video görüntülerinden yola çıkarak odanın sıcaklığını tahmin etmek.

Sınıflandırma problemleri

- Çalışan verisi kullanarak kişinin işten ayrılıp ayrılmayacağını tespit etmek
- Spam mail tespiti
- Fotoğrafta kedi mi var köpek mi var tespiti

Linear Regression

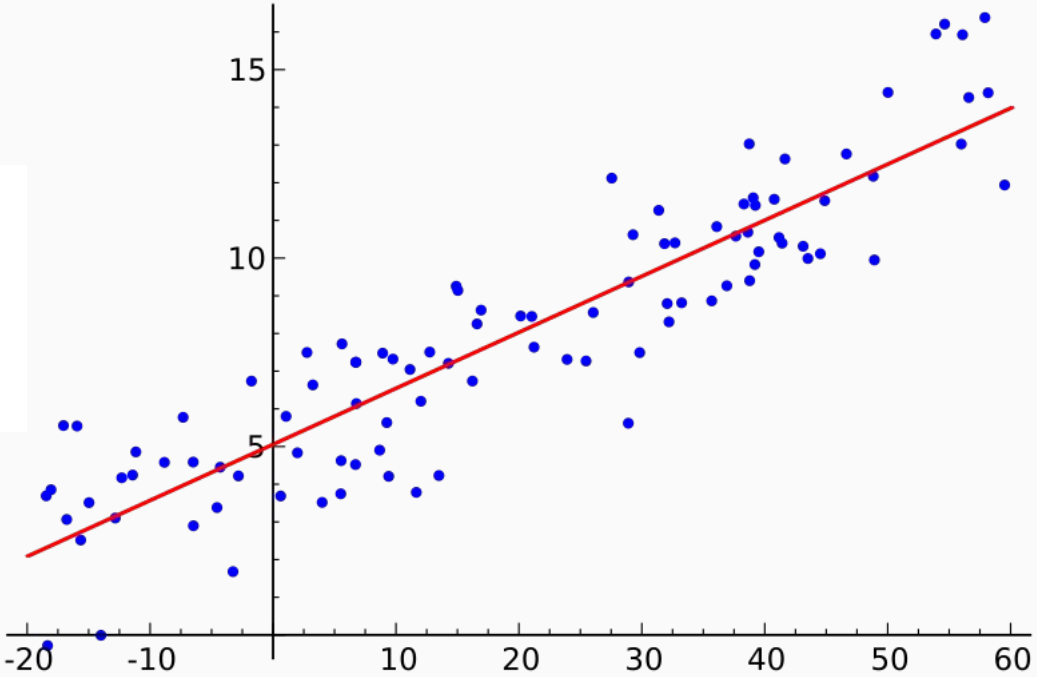
$$y(x) = w_0 + w_1 * x$$



Linear Regression

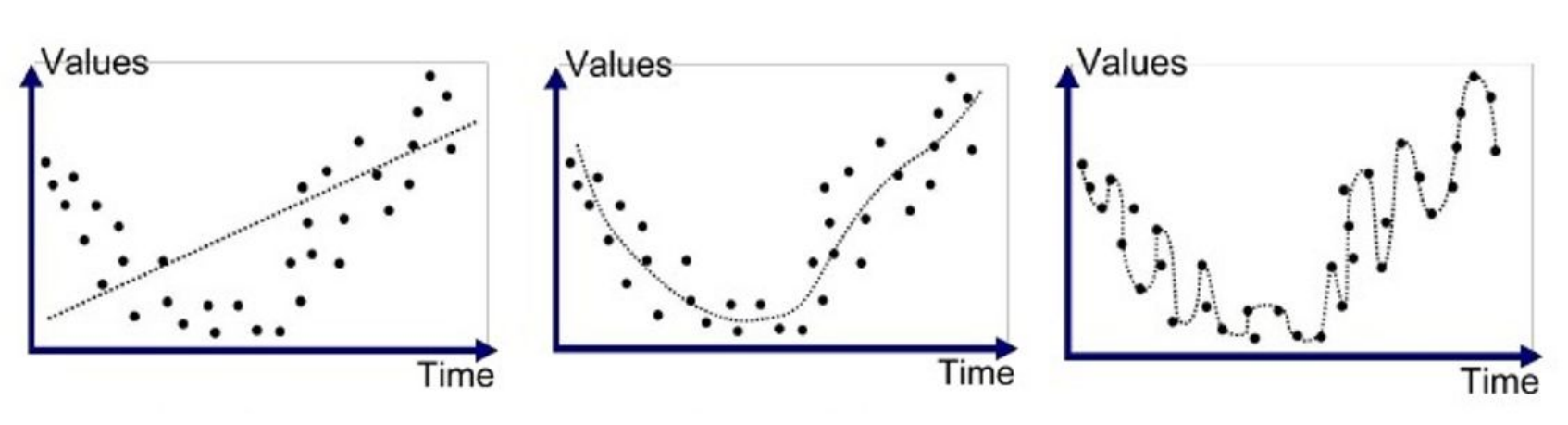
$$y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + \epsilon = \sum_{j=1}^{D} w_j x_j + \epsilon$$

- Outcome variable (real-valued labels)
- Coefficient vector
- Feature vector
- Residual error



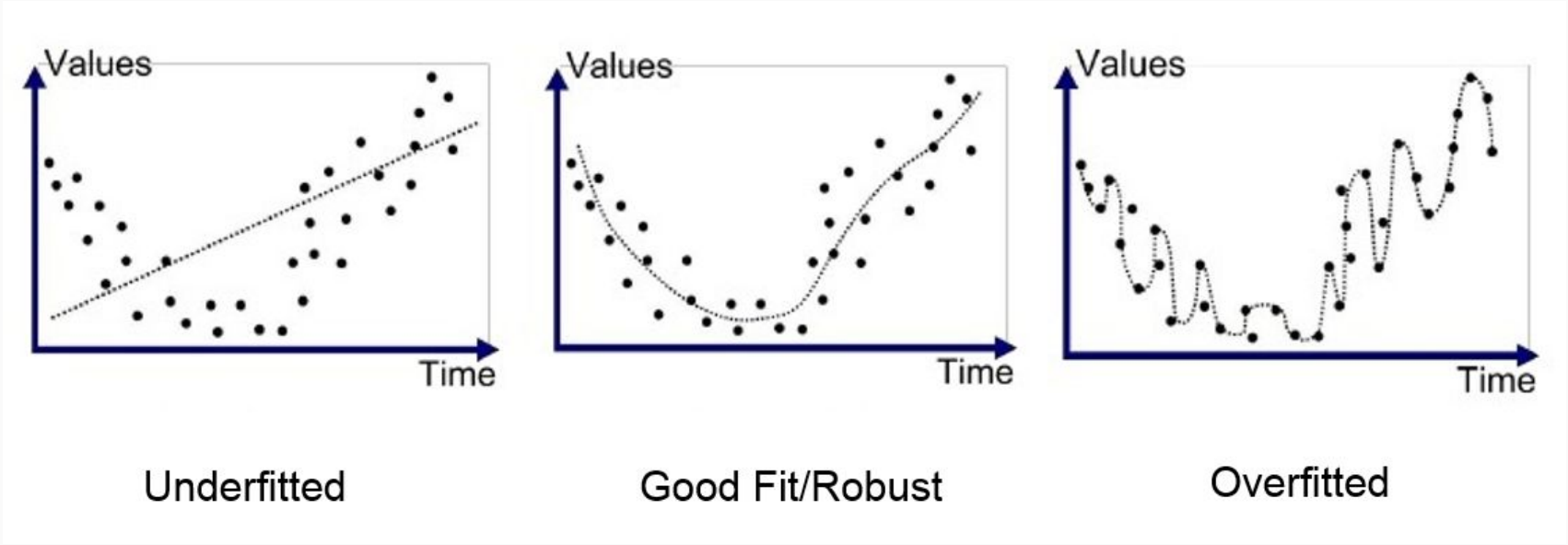
Karşılaşılabilecek Sıkıntılar

Overfitting / Underfitting: Hangi görseldekini tercih ederdiniz?



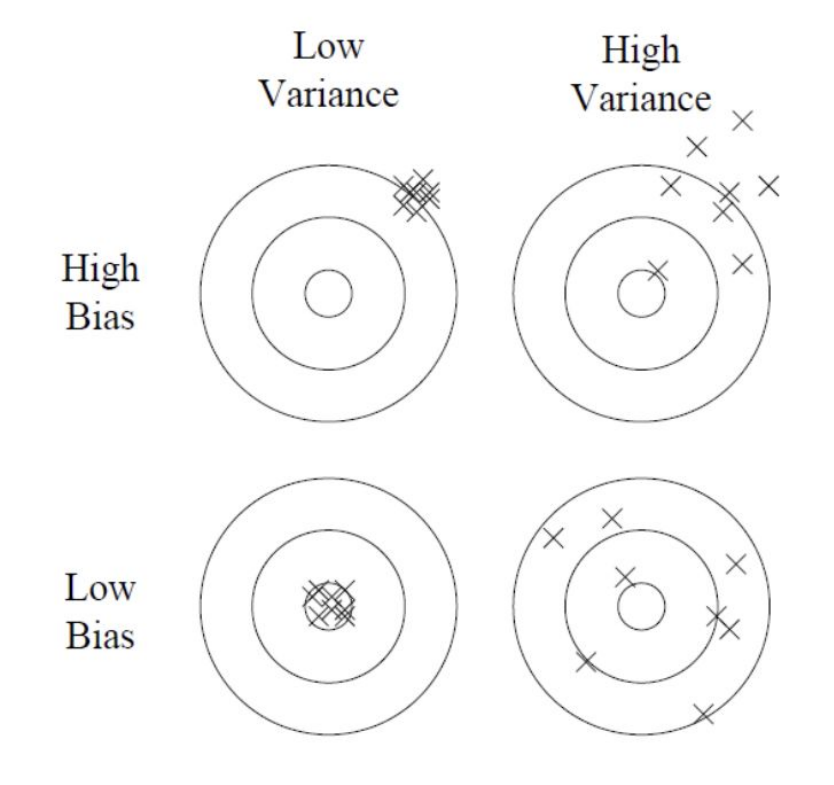
Karşılaşılabilecek Sıkıntılar

Overfitting / Underfitting: Hangi görseldekini tercih ederdiniz?



Karşılaşılabilecek Sıkıntılar

Bias / Variance

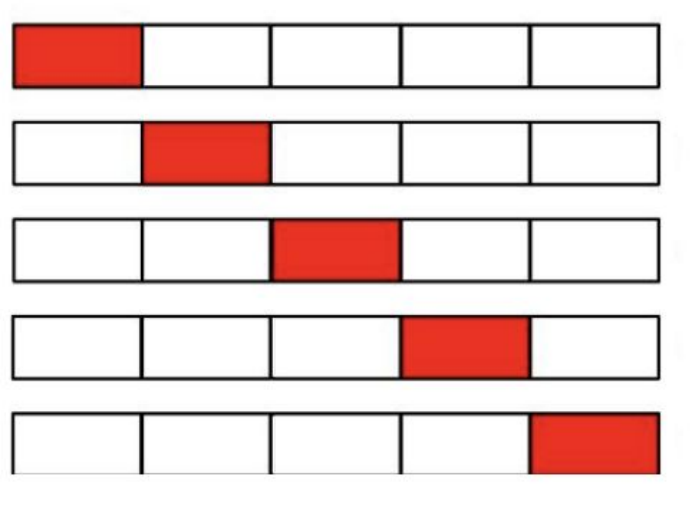


Model Seçimi ve Değerlendirmesi

Altın kural: ASLA test verisi ile model eğitme!

Test verisi farklı şekillerde seçilebilir.

- Random bir şekilde train-test diye iki sete bölmek.
- K-Fold Cross-Validation kullanmak.



Curse of Dimensionality

Büyük boyutun laneti:

- Veriye her bir yeni öznitelik eklendiğinde vektör uzayı iki katına çıkar.
- Vektör uzayı iki katına çıkarken verideki örnek sayısı bunu karşılayamazsa veri seyrekleşir, ve performans kaybı yaşanır.

Curse of Dimensionality

Büyük boyutun laneti:

- Veriye her bir yeni öznitelik eklendiğinde vektör uzayı iki katına çıkar.
- Vektör uzayı iki katına çıkarken verideki örnek sayısı bunu karşılayamazsa veri seyrekleşir, ve performans kaybı yaşanır.

Laneti Kırma Yöntemleri:

- **Feature Selection:** Veri içerisinden en değerli özniteliklerin tespit edilmesi ve bunlarla bir model eğitilmesi. Yapılması uzun süren maliyetli bir işlemdir.
- **Dimensionality Reduction:** Veri içerisindeki belli desenleri tespit ederek minimum bilgi kaybıyla aynı veriyi daha az öznitelikle temsil etme işlemidir.
  - Principal Component Analysis (PCA)
  - Singular Value Decomposition (SVD)

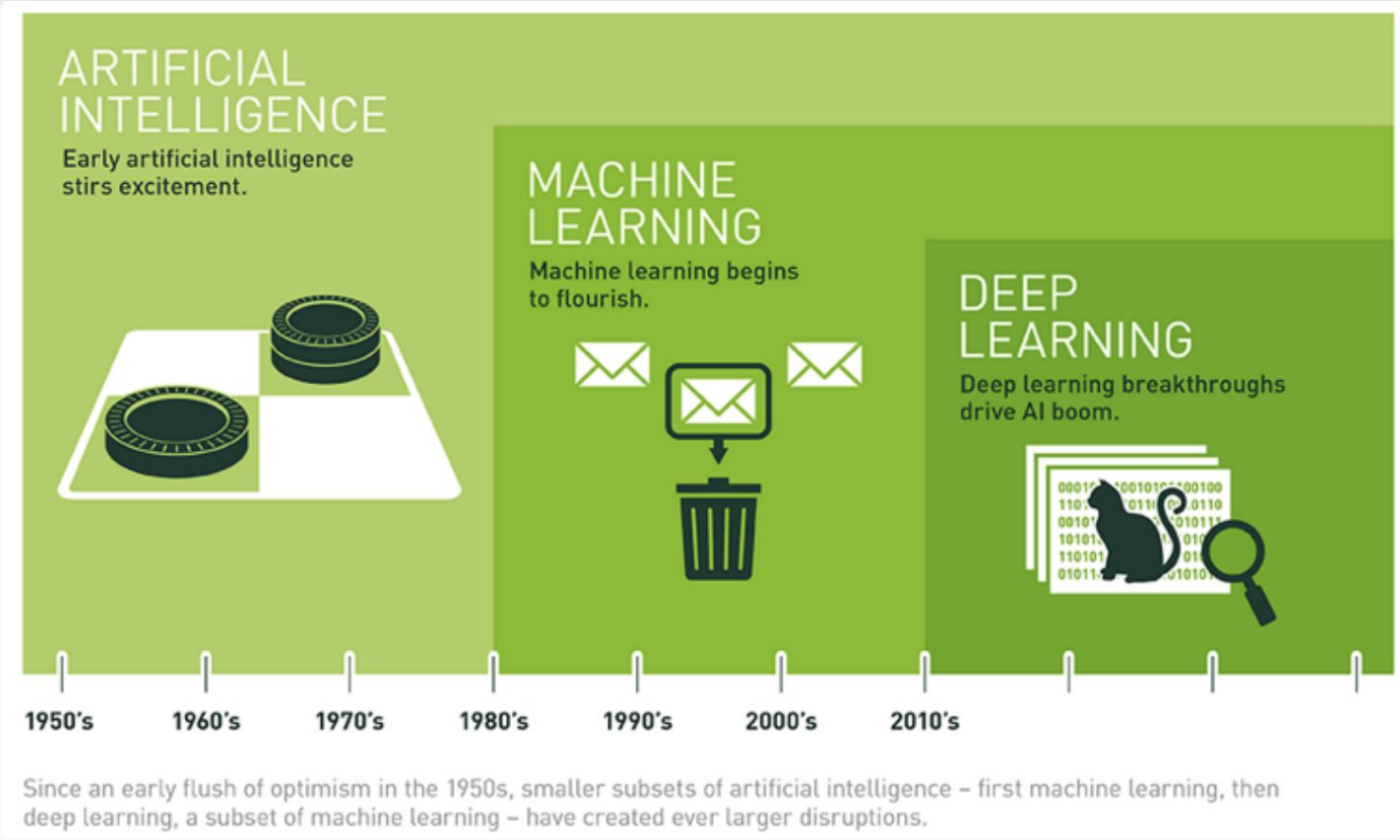
PCA

Mean

Principal Component 1

Principal Component 2

Yapay Zeka vs Makine Öğrenimi vs Derin Öğrenme



Kurs Kapsamı

Kurs ne içeriyor?

- Keşifçi Veri Analizi
  - Python ile veri analizi, veri görselleştirme
  - Veri içerisindeki ilişkilerin tespiti
- Veri Mühendisliği
  - Feature Extraction (Öznitelik Çıkarma Teknikleri)
  - Feature Selection (Öznitelik Seçme Teknikleri)
- Yapay Zeka Modelleri Eğitimi Pratik Bilgiler
  - Az teori ve çok pratik ile makine öğrenimi uygulamaları,
  - Farklı makine öğrenimi ve derin öğrenme modelleri hakkında genel bilgi
  - Model başarım metrikleri, model başarım ölçümü
  - Model optimizasyonu
- Model Serving
  - FastAPI aracılığıyla eğitilen modelin kullanıcılara sunulması

Kurs Kapsamı

Kurs ne içermiyor?

- Python dersleri
- Makine öğrenimi modelleriyle ilgili derinlemesine teorik bilgiler
- Unsupervised learning, semi-supervised learning, reinforcement learning
- CNN, RNN, LSTM, BERT, GPT gibi derin öğrenme modellerinin teknik anlatımları (genel bilgi olarak anlatılacak)

Q&A