

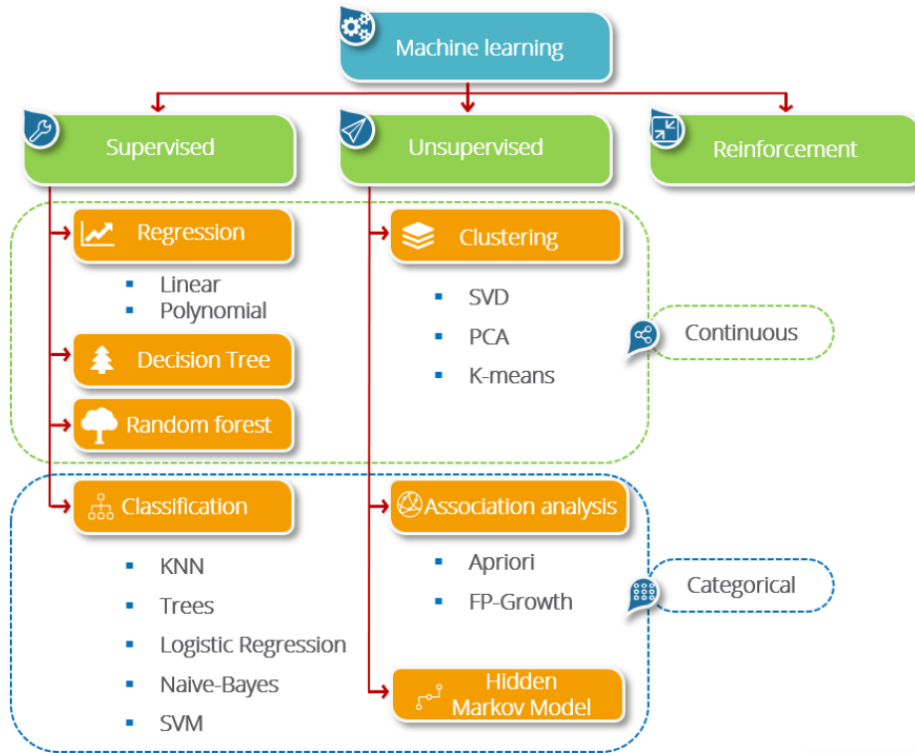
## K-NEAREST NEIGHBORS

### İçerik:

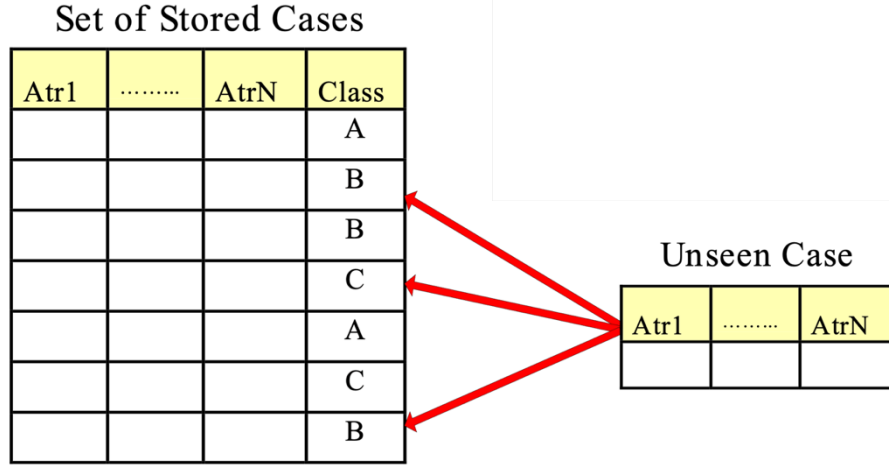
- K-En yakın komşu nedir?
- Çalışma mantığı nasıldır?
- Uzaklık Hesaplama Yöntemleri
- Kullanım alanları nerelerdir?
- Özellikleri
- Kaynakça

### K-En Yakın Komşu Nedir?

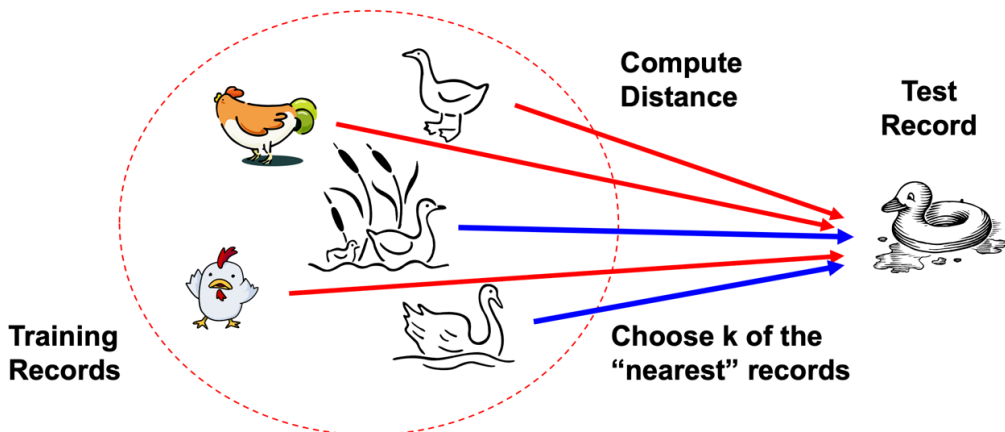
K-NN, denetimli makine öğrenmesi grubunun sınıflandırma yöntemlerinden biridir. Denetimli makine öğrenmesinde algoritma üzerinde bir öğretmen görevi üstleniyoruz diyebiliriz. Yani algoritmaya hangi sonuçları çıkaracağını önceden öğretiyoruz. Algoritma da kendisine sunulan gözlemi bizim ona daha önceden öğrettiğimiz verilerle karşılaştırarak uygun bir sınıfa atıyor. İşte K-en yakın komşu algoritması da bu mantığa dayalı çalışıyor. Aşağıdaki sınıflandırma basamaklarında da K-NN'nin sınıflandırmadaki konumunu görüyoruz.



Şimdi K-NN'ye kendi mantığına biraz daha inecek olursak elimizde bilinmeyen bir case var. Bu case'in özelliklerini daha önce öğrendiği sınıfların özellikleriyle karşılaştırıyor ve en yakın benzerlikteki sınıfa atıyor. Aşağıdaki şekilde bunu görüyoruz.

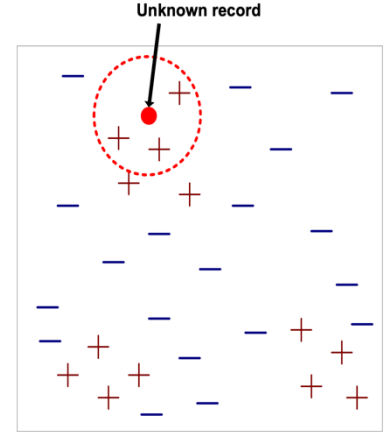


Biraz daha somut bir şeyin üstünden örnek vermek gerekirse aşağıdaki şekle bir bakalım. Burada bir test kaydımız var, bir ördek. Burada K-NN algoritması, daha önceden modele gösterdiğimiz kayıtlarla test kaydımızı karşılaştırıyor. Ördek gibi yürüyor ve ördek gibi ses çıkarıyorsa bu bir ördektir mantığıyla bu bir ördektir diyor.



K-NN sınıflandırma yaparken ‘en yakın’ k tane noktayı(en yakın komşular) kullanır ve bu sınıflandırmayı yaparken üç şeye ihtiyaç duyar.

- Etiketlenmiş kayıtlar kümesi
- Uzaklık metriği(kayıtlar arasındaki mesafeyi hesaplamak için) ve
- K değeri(alınacak en yakın komşuların sayısı)



Tüm eğitim verilerini ezberler ve yalnızca kaydın özelliklerini, eğitim örneklerinden biriyle tam olarak eşleyebilirse sınıflandırmayı başarır.

-- Bu durumda test kaydı, herhangi bir eğitim örneğiyle eşleşmediği için onu sınıflandıramayabilir.

### Çalışma Mantığı Nasıldır?

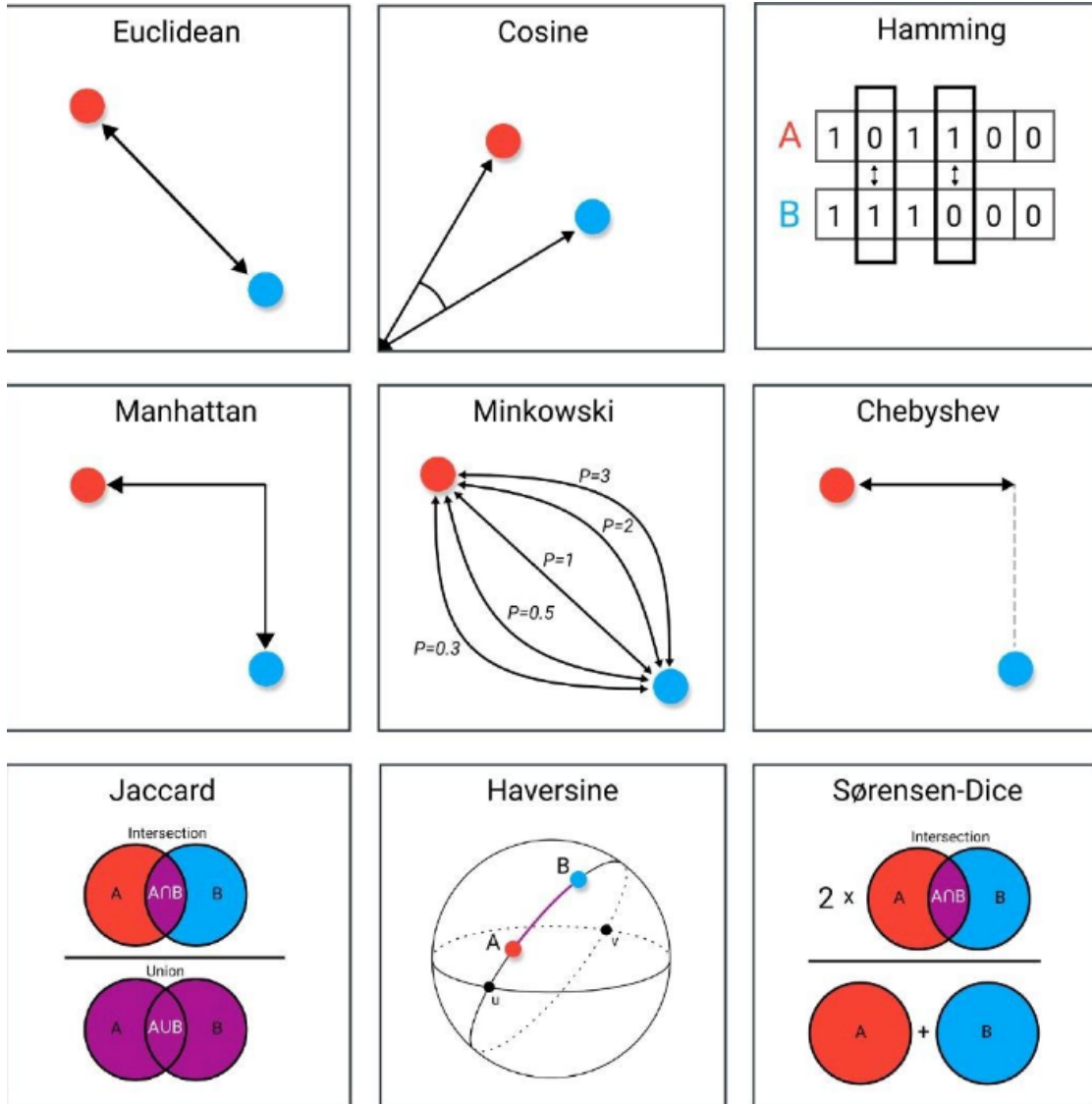
Yukarıda bahsettiğimi şöyle bir toparlarsak çalışma mantığının adımları şu şekildedir:

- Verinin, diğer verilerle (yani komşularla) olan benzerliğini hesapla.
- Komşular arasından en yakın K tanesini seç.
- Bu K komşunun her biri için, hangi sınıfta yer aldığına bak.
- Örneğin A sınıfı, K komşu içerisinde en çok temsil edilen sınıfsa, elimizdeki veriyi o sınıfa yaz.



## Uzaklık Hesaplama Yöntemleri

Bunun için birçok yöntem var. Aşağıdaki resimde bu yöntemleri görüyoruz.



En çok kullanılan üç yöntemden biraz daha detaylı bahsedeyim.

Euclidean (Öklidyen) Uzaklık Hesaplaması; Öklid Uzayı  $R^n$  içinde, iki nokta arası verilen mesafe hesaplaması olarak ifade edilmektedir.

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Manhattan Uzaklık Hesaplaması;  $\mathbb{R}^n$  içinde (uyarlanabilir) eksenler boyunca dik açılarda ölçülen iki nokta arasındaki mesafe hesaplaması olarak ifade edilmektedir.

$$d_1(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum_{i=1}^n |p_i - q_i|$$

Minkowski Uzaklık Hesaplaması;  $\mathbb{R}^n$  içinde (uyarlanabilir) normlu kullanılan bir metrik\* mesafe hesaplaması olarak ifade edilmektedir.

$$P = (x_1, x_2, \dots, x_n) \text{ and } Q = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$$

$$\left( \sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

### Kullanım Alanları Nerelerdir?

K-NN günlük hayatımızda birçok alanda kullanılıyor. Bunlara birkaç örnek vermek gerekirse şu şekilde sıralayabiliriz:

- Bankacılık sektöründe
  - Kredi ve Risk Analizleri
  - İşletme İflas Tahminlemeleri
  - Borsa Analizleri (Alım-Satım İşlemleri) gibi işlemlerde
- Ekolojik Eylem Analizleri
- Biyolojik Analizler
- Kompleks Metin Madenciliği Analizleri
- Farklılaşmış Hastalık Analizleri (İleri Seviye Kanser/Tümör)
- Benzerliğe Dayalı Kayıp Veri Analizleri (Kompleks Seviye)

gibi birçok alanda kullanıyor.

## Özellikleri

- k-NN sınıflandırıcılar, modeli açıkça oluşturmadıkları içintembel öğrencilerdir (lazy learners )
- Bilinmeyen kayıtları sınıflandırmak nispeten maliyetlidir (Karar ağacı indüksiyonu ve kural tabanlı sistemler gibi istekli (eager) öğrencilerin aksine)
- Keyfi şekillendirilmiş karar sınırları üretebilir (arbitrarily shaped decision boundaries)
- Kararlar yerel bilgilere dayandığından, değişken etkileşimlerini yönetmek kolaydır
- Sınıflandırma kararları yerel olarak verildiği için, en yakın komşu sınıflandırıcılar (küçük k değerleri için) gürültüye oldukça duyarlıdır.
- Doğru yakınlık ölçüsünün seçimi önemlidir
- Gereksiz ve fazlalık öznelilikler sorun yaratabilir
- Eksik özelliklerin üstesinden gelmek zordur

## Kaynakça

- Introduction to Data Mining , 2nd Edition by Tan, Steinbach, Karpatne, Kumar
- <https://mervetatlidil.medium.com/makine-öğrenmesi-machine-learning-8960166d36d8>
- <https://arslanev.medium.com/makine-öğrenmesi-knn-k-nearest-neighbors-algoritması-bdfb688d7c5f>
- <https://medium.com/machine-learning-türkiye/localized-regression-knn-with-local-regression-7b4d302adb85>
- [https://miracozturk.com/python-ile-siniflandirma-analizleri-knn-k-nearest-neighbours-k-en-yakin-komsu-algoritmasi/#Euclidean\\_Oklidyen\\_Uzaklik\\_Hesaplama](https://miracozturk.com/python-ile-siniflandirma-analizleri-knn-k-nearest-neighbours-k-en-yakin-komsu-algoritmasi/#Euclidean_Oklidyen_Uzaklik_Hesaplama)
- <http://bilgislem.com/basit-bir-yapay-zeka-algoritmasi-kodlayalim-knn-algoritmasi/>
- [https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm)
- <https://medium.com/@k.ulgen90/makine-öğrenimi-bölüm-2-6d6d120a18e1>
- <https://miracozturk.com/python-ile-siniflandirma-analizleri-knn-k-nearest-neighbours-k-en-yakin-komsu-algoritmasi/>

Hazırlayan: Melike Yoğurtcu