

KNN

ALGORİTMASI İLE KANSER TESPİTİ

KNN ALGORİTMASI NEDİR?

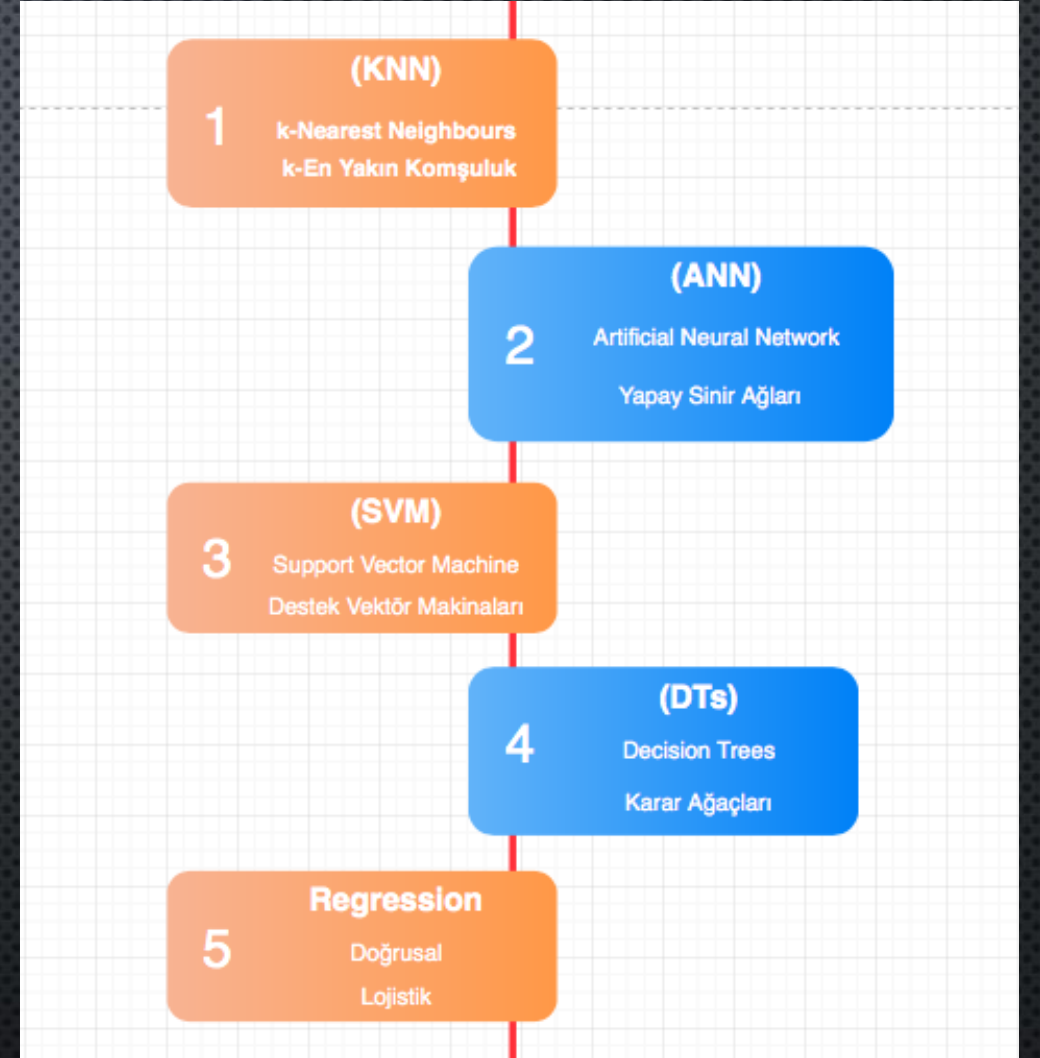
NASIL KULLANILIR?

ADIMLARI NELERDİR?

İLHAN ERSOY-HASAN DEMİR

KNN ALGORİTMASI NEDİR?

K-en yakın komşuluk (KNN) algoritması, uygulaması kolay gözetimli öğrenme algoritmalarındandır. Hem sınıflandırma hem de regresyon problemlerinin çözümünde kullanılıyor olmakla birlikte, endüstride çoğunlukla sınıflandırma problemlerinin çözümünde kullanılmaktadır.

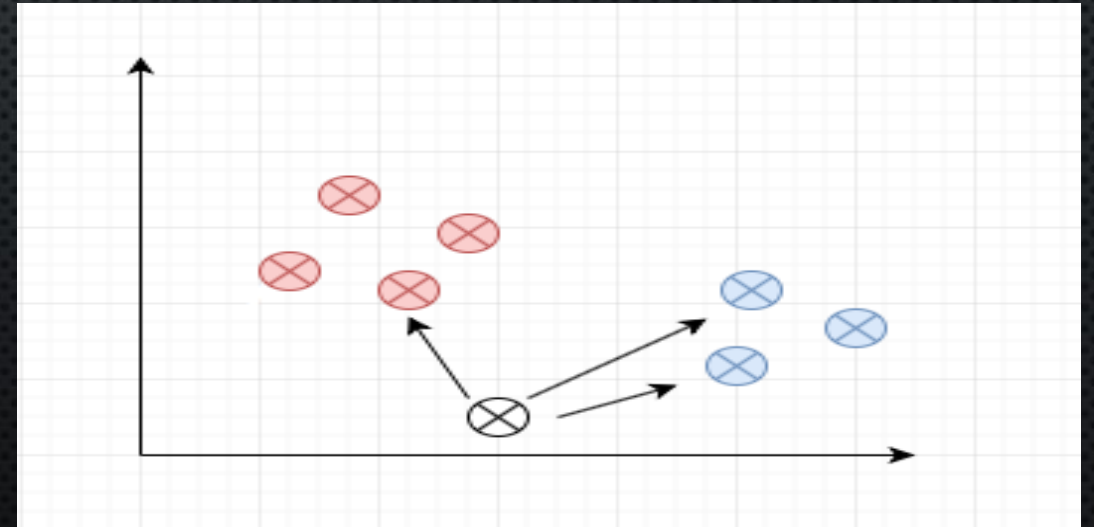


- KNN ALGORİTMALARI, 1967 YILINDA T. M. COVER VE P. E. HART TARAFINDAN ÖNERİLMİŞTİR. ALGORİTMA, SINIFLARI BELLİ OLAN BİR ÖRNEK KÜMESİNDEKİ VERİLERDEN YARARLANILARAK KULLANILMAKTADIR. ÖRNEK VERİ SETİNE KATILACAK OLAN YENİ VERİNİN, MEVCUT VERİLERE GÖRE UZAKLIĞI HESAPLANIP, K SAYIDA YAKIN KOMŞULUĞUNA BAKILIR. UZAKLIK HESAPLARI İÇİN GENELDE 3 TİP UZAKLIK FONKSİYONU KULLANILMAKTADIR:
- “EUCLİDEAN” UZAKLIK
- “MANHATTAN” UZAKLIK
- “MINKOWSKI” UZAKLIĞI'DIR.

KNN; eski, basit ve gürültülü eğitim verilerine karşı dirençli olması sebebiyle en popüler makine öğrenme algoritmalarından biridir. Fakat bunun yanında dezavantajı da mevcuttur. Örneğin, uzaklık hesabı yaparken bütün durumları sakladığından, büyük veriler için kullanıldığında çok sayıda bellek alanına gereksinim duymaktadır.

Distance functions

Euclidean	$\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$
Manhattan	$\sum_{i=1}^k x_i - y_i $
Minkowski	$\left(\sum_{i=1}^k (x_i - y_i)^q \right)^{1/q}$



KNN ALGORİTMASININ ADIMLARI:

İLK OLARAK K PARAMETRESİ BELİRLENİR. BU PARAMETRE VERİLEN BİR NOKTAYA EN YAKIN KOMŞULARIN SAYISIDIR. ÖRNEĞİN: $K=3$ OLSUN. BU DURUMDA EN YAKIN 3 KOMŞUYA GÖRE SINIFLANDIRMA YAPILACAKTIR.

ÖRNEK VERİ SETİNE KATILACAK OLAN YENİ VERİNİN, MEVCUT VERİLERE GÖRE UZAKLIĞI TEK TEK HESAPLANIR. İLGİLİ UZAKLIK FONKSİYONLARI YARDIMIYLA.

İLGİLİ UZAKLILARDAN EN YAKIN K KOMŞU ELE ALINIR. ÖZNİTELİK DEĞERLERİNE GÖRE K KOMŞU VEYA KOMŞULARIN SINIFINA ATANIR.

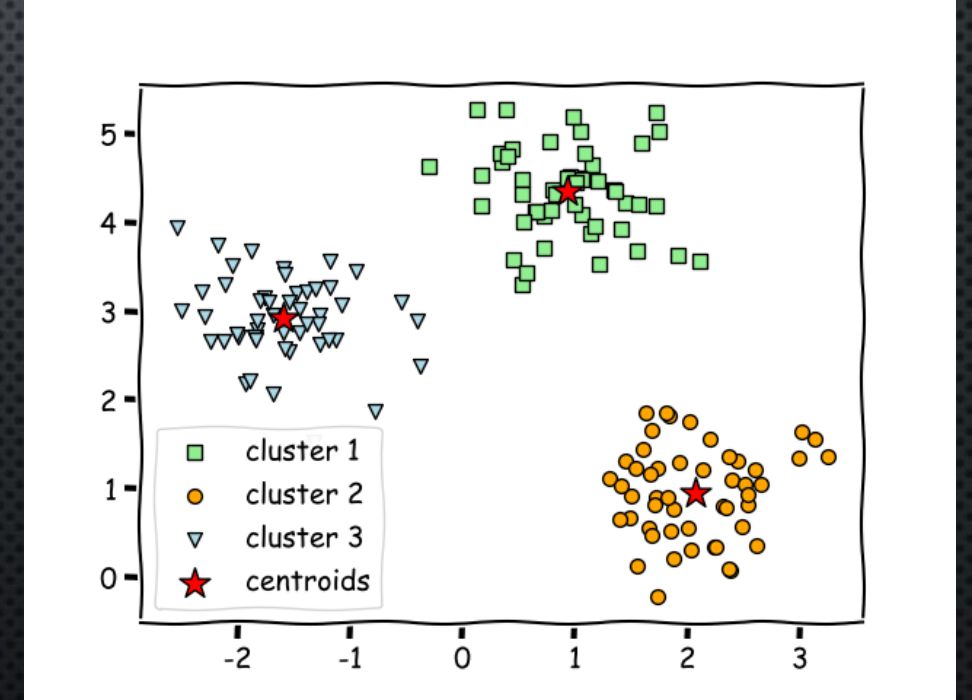
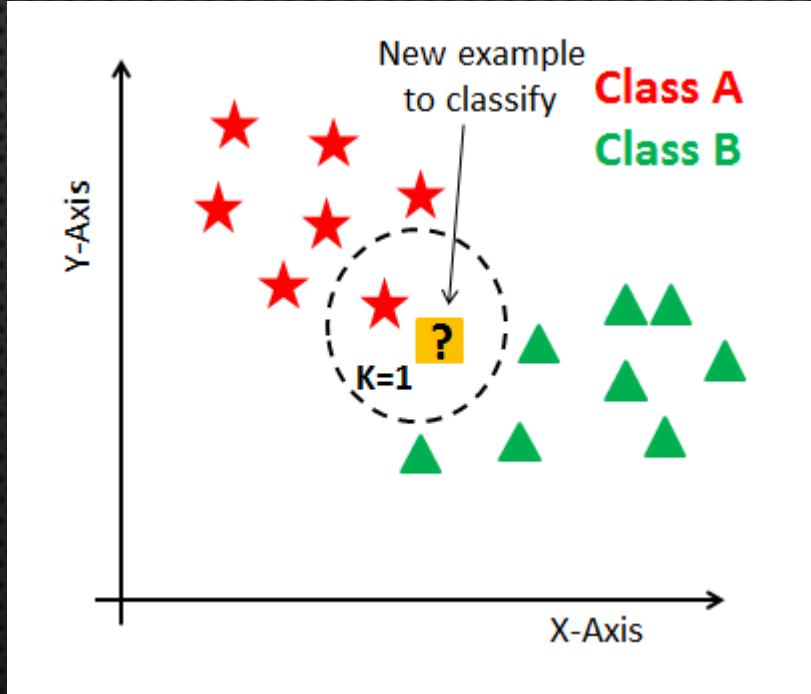
SEÇİLEN SINIF, TAHMİN EDİLMESİ BEKLENEN GÖZLEM DEĞERİNİN SINIFI OLARAK KABUL EDİLİR. YANI YENİ VERİ ETİKETLENMİŞ (LABEL) OLUR.



KNN İLE GÖĞÜS KANSERİ TESPIİTİ

- PROJENİN AMACI

KANSER HASTASI OLAN BİREYLERİN VERİLERİNE BAKILARAK BAŞKA BİR HASTANIN KANSER OLUP OLMADIĞINI EN YAKIN DEĞER ALGORİTMASI İLE TEST ETMEK.



PEKİ NASIL ÇALIŞIYOR?

- YARATTIĞIMIZ ALGORİTMA KANSERLİ VEYA KANSER OLMAYAN BİR HASTANIN VERİLERİNİN DİĞER TÜM HASTALARA OLAN UZAKLIKLARINI HESAPLIYOR VE KÜÇÜKTEN BÜYÜĞE SIRAYA DİZİYOR.
- BU DİZİM İŞLEMİNDEN SONRA $K = 5$ OLACAK ŞEKİLDE KENDİSİNE EN YAKIN 5 VERİNİN SINIFLARINA BAKIYOR (KANSER OLUP OLMADIĞINA) VE EN ÇOK HANGİ SINIFI İÇERİYORSA SONUCA GÖRE BİR TAHMİNDE BULUNUYOR
- ÖRNEK OLARAK BURADA TEST VERİSİNDE SATIR 82 DE Kİ VERİNİN EĞİTİMDEKİ 600 VERİ İÇİNDE KENDİSİNE EN YAKIN 5 VERİNİN HASTA OLUP OLMADIĞINA BAKIYOR VE SAYICA FAZLA OLAN DURUMA GÖRE BU ÇIKTIYI VERİYOR.

```
TEST SATIRI:82
```

```
82 numarali test verisine en yakin komsular:
```

```
Kiyaslanan veriye olan uzaklik:29 Egitim Verisinin Hastalik Durumu:4  
Kiyaslanan veriye olan uzaklik:37 Egitim Verisinin Hastalik Durumu:4  
Kiyaslanan veriye olan uzaklik:40 Egitim Verisinin Hastalik Durumu:4  
Kiyaslanan veriye olan uzaklik:43 Egitim Verisinin Hastalik Durumu:4  
Kiyaslanan veriye olan uzaklik:44 Egitim Verisinin Hastalik Durumu:4
```

```
Hastalikli veri:5
```

```
Saglikli veri:0
```

```
Hasta!
```

Bunu Öklid uzaklığı ile gerçekleştiriyor.

BUNU NASIL YAPIYOR?

*test.txt - Not Defteri

Dosya	Düzen	Biçim	Görünüm	Yardım
SATIR 0 => 1041043,4,1,3,1,2,1,2,1,1,2				
HASTA İD uzaklığına bakılan veriler				

Testte ki bir hastanın tüm verilerini Eğitimdeki hastaların her birine uzaklığını buluyor ve sıraya diziyor.

Örnek olarak testteki SATIR 0 'a en yakın 3 değer SATIR 1 (sağlıklı '2'), SATIR 4 (sağlıklı '2') ve SATIR 6 (sağlıklı '2') olsun.

Bu durumda Satır 0'a konulacak teşhis 'Sağlıklı' olacaktır.

(3 SAĞLIKLI VERİ - 0 HASTA VERİ)

*eğitim.txt - Not Defteri

Dosya	Düzen	Biçim	Görünüm	Yardım
Satır 0 => 1000025,5,1,1,1,2,1,3,1,1,2				
Satır 1 =>1002945,5,4,4,5,7,10,3,2,1,2				
Satır 2 =>1015425,3,1,1,1,2,2,3,1,1,2				
Satır 3 =>1016277,6,8,8,1,3,4,3,7,1,2				
Satır 4 =>1017023,4,1,1,3,2,1,3,1,1,2				
Satır 5 =>1017122,8,10,10,8,7,10,9,7,1,4				
Satır 6 =>1018099,1,1,1,1,2,10,3,1,1,2				
Satır 7 =>1018561,2,1,2,1,2,1,3,1,1,2				
Satır 8 =>1033078,2,1,1,1,2,1,1,1,5,2				
.				
.				
.				
.				
.				
Satır 599 =>1031608,2,1,1,1,1,1,2,1,1,2				