

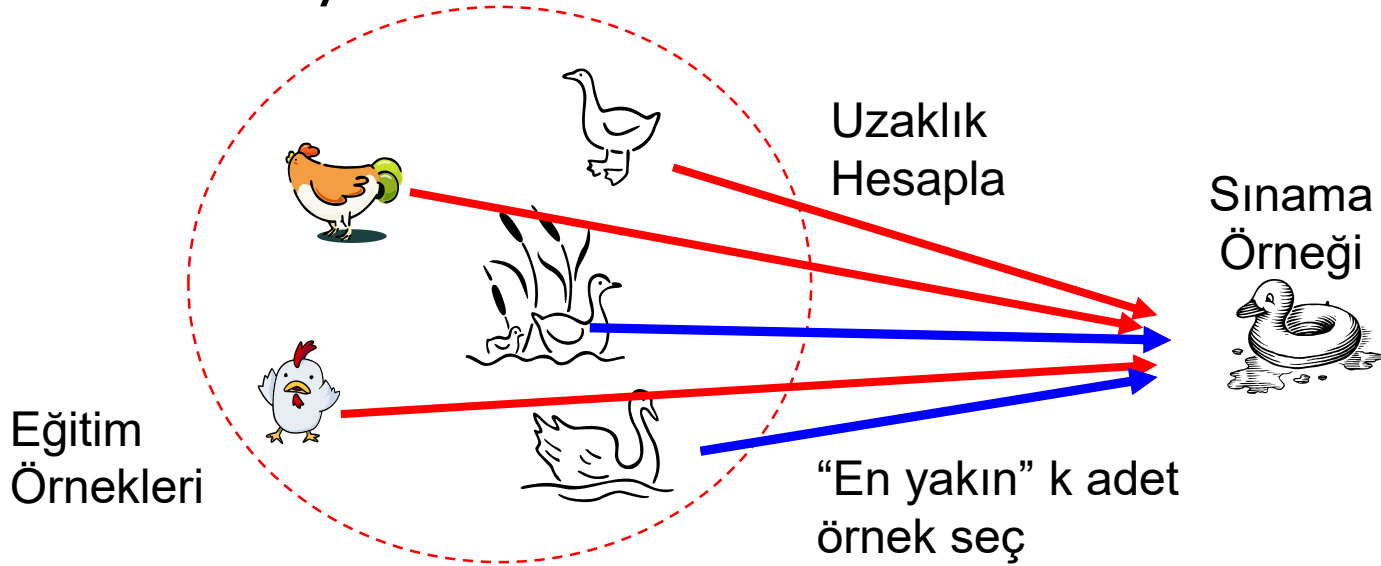


Örnek Tabanlı Yöntemler

- Örnek tabanlı sınıflandırma:
 - Öğrenme kümesi saklanır.
 - Sınıflandırılacak yeni bir örnek geldiğinde öğrenme kümesi sınıf etiketini öngörmek için kullanılır.
- Yöntemler
 - k-en yakın komşu yöntemi

K-En Yakın Komşu

- Temel yaklaşım: Sınıflandırılmak istenen örneğe en yakın örnekleri bul.
 - Örnek: ördek gibi yürüyor, ördek gibi bağıriyor
 - => büyük olasılıkla ördek





K-En Yakın Komşu

- Sınıfları belli olan bir örnek kümesindeki gözlem değerlerinden yararlanarak örneğe katılacak yeni bir gözlemin hangi sınıfa ait olduğunu belirlemek amacıyla kullanılır.
- Yöntem örnek kümedeki gözlemlerin her birinin sonradan belirlenen bir gözlem değerine olan uzaklıklarının hesaplanması ve **en küçük uzaklığa sahip k sayıda gözlemin** seçilmesi esasına dayanır.



K-En Yakın Komşu

- k en yakın komşuluğunda temel düşünce “komşunun yaptığı gibi yap” tır. Eğer belirli bir kişinin davranışı tahmin edilmek isteniyorsa, veri uzayında o kişiye yakın, kişinin davranışlarına bakılır.
- Bu kişilere ait davranışlarının ortalaması hesaplanır ve bu ortalama belirlenen kişi için tahmin olur.



K-En Yakın Komşu

- Kayıtlar, bir veri uzayındaki noktalar olarak düşünülürse, birbirine yakın olan kayıtlar, birbirinin civarında (yakın komşusu) olur.
- k-en yakın komşuluğunda, k parametresi araştırılan komşuların sayısıdır. 5-yakın komşuluğunda 5 kişiye, 1-yakın komşuluğunda 1 kişiye bakılır.



K-En Yakın Komşu

- Uzaklık hesaplamaları için 3 uzaklık fonksiyonu kullanılmaktadır.

Euclidean

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

Manhattan

$$\sum_{i=1}^k |x_i - y_i|$$

Minkowski

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

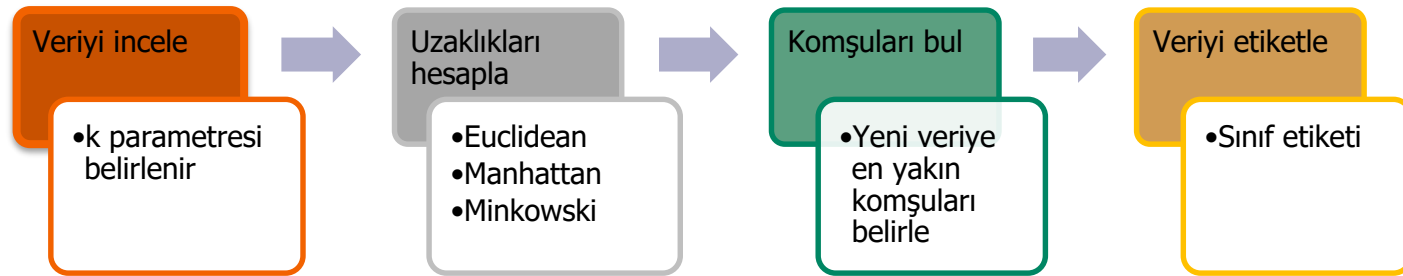


K-En Yakın Komşu

- Genellikle Öklid uzaklık formülü kullanılır.



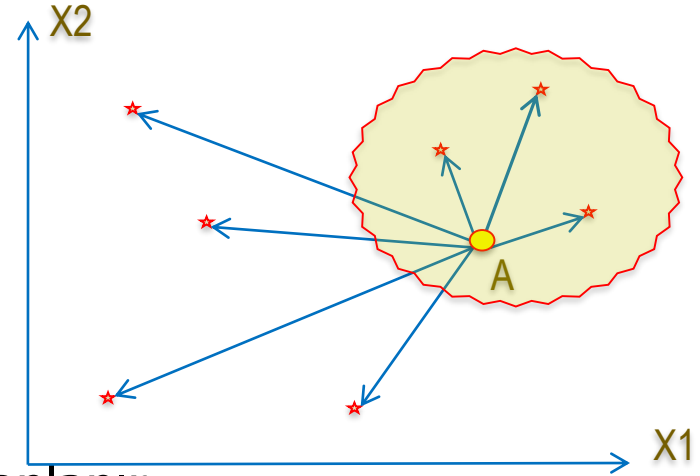
K-En Yakın Komşu



K-En Yakın Komşu

- k-en yakın komşu algoritmasının adımları:

- k parametresi belirlenir.
- Komşuluklara ait uzaklıklar hesaplanır.
- Hesaplanan uzaklıklara göre satırlar sıralanarak en küçük k tanesi belirlenir.
- Belirlenen satırların hangi sınıfa ait olduğu bulunarak tekrarlanan sınıf değeri seçilir.
- Seçilen sınıf, tahmin edilmesi beklenen gözlem değerinin sınıfı olarak kabul edilir.



K-En Yakın Komşu - Örnek

- Aşağıdaki tablo X, Y gözlem değerlerinden ve Z sınıf değerlerinden oluşmaktadır. Bu gözlem değerleriyle yola çıkarak yeni verilen gözlem değerinin hangi sınıfa ait olduğunu k- en yakın komşu yöntemiyle bulalım.

Yeni gözlem değeri X=8, Y=4;

X	Y	Z
2	4	Negatif
3	6	Pozitif
3	4	Pozitif
4	10	Negatif
5	8	Negatif
6	3	Pozitif
7	9	Pozitif
9	7	Negatif
11	7	Negatif
10	2	Negatif

1. Adım :

k= 4 için işlem yapalım.

(Problem için (8,4) noktasına en yakın komşu değerleri arayalım.)

K-En Yakın Komşu - Örnek

2. Adım :

Öklit bağıntısına göre her bir gözlem değeri için uzaklıkları hesaplayalım.

Öklid uzaklık formülü ile (8,4) noktasının tüm gözlem değerleri ile arasındaki uzaklıkları hesaplayalım.

X	Y	Z
2	4	Negatif
3	6	Pozitif
3	4	Pozitif
4	10	Negatif
5	8	Negatif
6	3	Pozitif
7	9	Pozitif
9	7	Negatif
11	7	Negatif
10	2	Negatif

$$d(i, j) = \sqrt{(2-8)^2 + (4-4)^2} = 6.00$$

$$d(i, j) = \sqrt{(3-8)^2 + (6-4)^2} = 5.39$$

$$d(i, j) = \sqrt{(3-8)^2 + (4-4)^2} = 5.00$$

...

Hesaplanan değerler farklı bir tablo üzerinde gösterilirse...

K-En Yakın Komşu - Örnek

Gözlenen değerlerin (8,4) noktasına olan uzaklığı...

X	Y	Uzaklık
2	4	6.00
3	6	5.39
3	4	5.00
4	10	7.21
5	8	5.00
6	3	2.24
7	9	5.10
9	7	3.16
11	7	4.24
10	2	2.83

Uzaklık değerlerine göre $k=4$ komşu değerlerinin belirlenmesi

3. Adım :

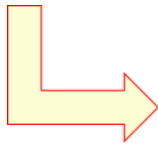
En küçük uzaklıkların belirlenmesi için satırlar sıralanarak en küçük $k=4$ tanesi belirleniyor.

Belirlenen dört nokta (8,4) noktasına en yakın değerlerdir.

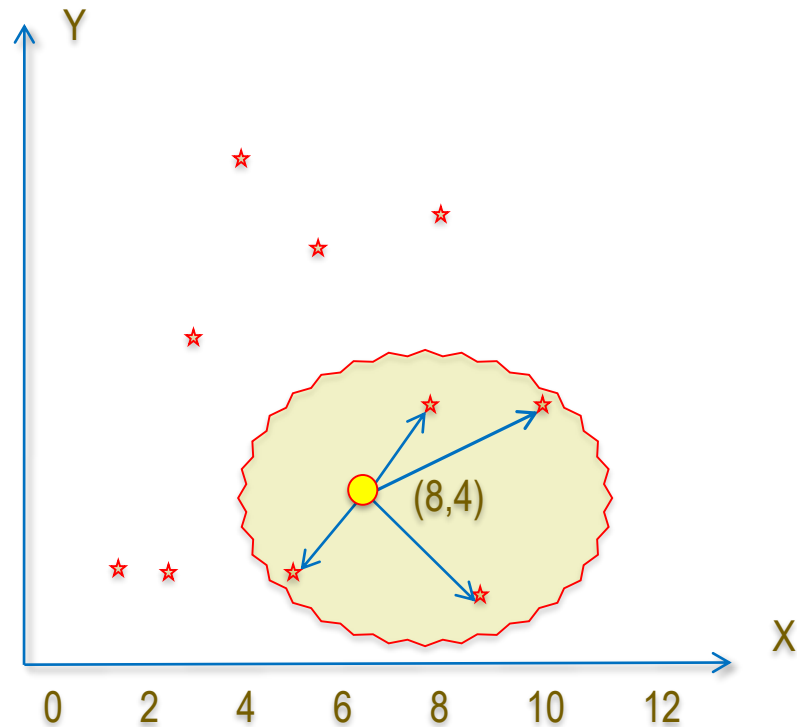
X	Y	Uzaklık	(En Küçük) Sıra
2	4	6.00	9
3	6	5.39	8
3	4	5.00	6
4	10	7.21	10
5	8	5.00	5
6	3	2.24	1
7	9	5.10	7
9	7	3.16	3
11	7	4.24	4
10	2	2.83	2

K-En Yakın Komşu - Örnek

X	Y	Uzaklık	(En Küçük) Sıra
2	4	6.00	9
3	6	5.39	8
3	4	5.00	6
4	10	7.21	10
5	8	5.00	5
6	3	2.24	1
7	9	5.10	7
9	7	3.16	3
11	7	4.24	4
10	2	2.83	2



(8,4) Noktasına komşu olan en yakın dört gözlenen değer koordinat sistemi üzerindeki gösterimi ile ...



K-En Yakın Komşu - Örnek

4. Adım : En küçük satırlara ilişkin sınıfların belirlenmesi işlemi gözlem değerlerinin içinde **hangi değerin baskın olduğuna göre** karar verilir.

X	Y	Uzaklık	(En Küçük) Sıra	Z
2	4	6.00	9	Negatif
3	6	5.39	8	Pozitif
3	4	5.00	6	Pozitif
4	10	7.21	10	Negatif
5	8	5.00	5	Negatif
6	3	2.24	1	Pozitif
7	9	5.10	7	Pozitif
9	7	3.16	3	Negatif
11	7	4.24	4	Negatif
10	2	2.83	2	Negatif

5. Adım :

Gözlem değerlerin içinde **bir pozitif** ve **üç negatif** değer olduğundan **(8,4)** noktasının sınıfı **negatif** olarak belirlenir.

(8,4) noktasının

Sınıfı Negatif
olarak belirlenir.

K-En Yakın Komşu - Örnek

- Üç gözlem değerine sahip tablonun sınıf değeri Q sütunu tarafından temsil edilmektedir. (7,8,5) noktasının hangi sınıf değerine sahip olduğunu tespit etmeye çalışalım.

X	Y	Z	Q
10	5	19	DOĞRU
8	2	4	YANLIŞ
18	16	6	YANLIŞ
12	15	8	DOĞRU
3	15	15	DOĞRU

Tabloda verilen değerleri dönüşüm ile min-max normalleştirme yöntemi ile gözlem değerlerini tekrar elde etmek istersek...

(0,1) aralığına göre dönüştürmek için aşağıdaki bağıntı üzerinden işlem yaparsak ;



$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

K-En Yakın Komşu - Örnek

Bağıntıya gerekli olan değerleri tablo halinde gösterirsek



	X	Y	Z
X_{\min}	3	2	4
X_{\max}	18	16	19

İlk satırda **X** için X^* hesaplaması;

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} = \frac{10 - 3}{18 - 3} = 0.47$$

Benzer biçimde **Y** için ilk satır değerinin hesaplanması;

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} = \frac{5 - 2}{16 - 2} = 0.21$$

Aynı şekilde **Z** için ilk satır değerinin hesaplanması;

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} = \frac{19 - 4}{19 - 4} = 1$$

X	Y	Z	Q
10	5	19	DOĞRU
8	2	4	YANLIŞ
18	16	6	YANLIŞ
12	15	8	DOĞRU
3	15	15	DOĞRU

Benzer şekilde tablonun kalan değerleri de hesaplanarak tabloya aktarılırsa



K-En Yakın Komşu - Örnek

Dönüştürülmüş gözlem değerleri;

X	Y	Z	Q
0.47	0.21	1.00	DOĞRU
0.33	0.00	0.00	YANLIŞ
1.00	1.00	0.13	YANLIŞ
0.60	0.93	0.27	DOĞRU
0.00	0.93	0.73	DOĞRU

Sınıflandırma işlemine alınacak (7, 8, 5)

değerleride aynı dönüşüm formülüyle yeni değerlerine kavuşturulursa;

Yeni değerler (0.26, 0.43, 0.07) şeklinde elde edilir.

k-en yakın algoritması uygulanırsa, **k=3** alınarak öklid bağıntısına göre uzaklıklar hesaplandığında, değerleri tablo halinde görüntülersek;

X	Y	Z	Uzaklık
0.47	0.21	1.00	0.98
0.33	0.00	0.00	0.44
1.00	1.00	0.13	0.93
0.60	0.93	0.27	0.63
0.00	0.93	0.73	0.87

K-En Yakın Komşu - Örnek

En küçük uzaklıkların belirlenmesi: Satırlar sıralanarak en küçük k=3 tanesi tespit edilir.

X	Y	Z	Uzaklık	En küçük sıra
0.47	0.21	1.00	0.98	5
0.33	0.00	0.00	0.44	1 / YANLIŞ
1.00	1.00	0.13	0.93	4
0.60	0.93	0.27	0.63	2 / DOĞRU
0.00	0.93	0.73	0.87	3 / DOĞRU

Seçilen gözlem değerleri arasında DOĞRU sayısı baskın olduğundan (0.26,0.43,0.07) noktasının sınıfı **DOĞRU** olarak belirlenmiş olur.