

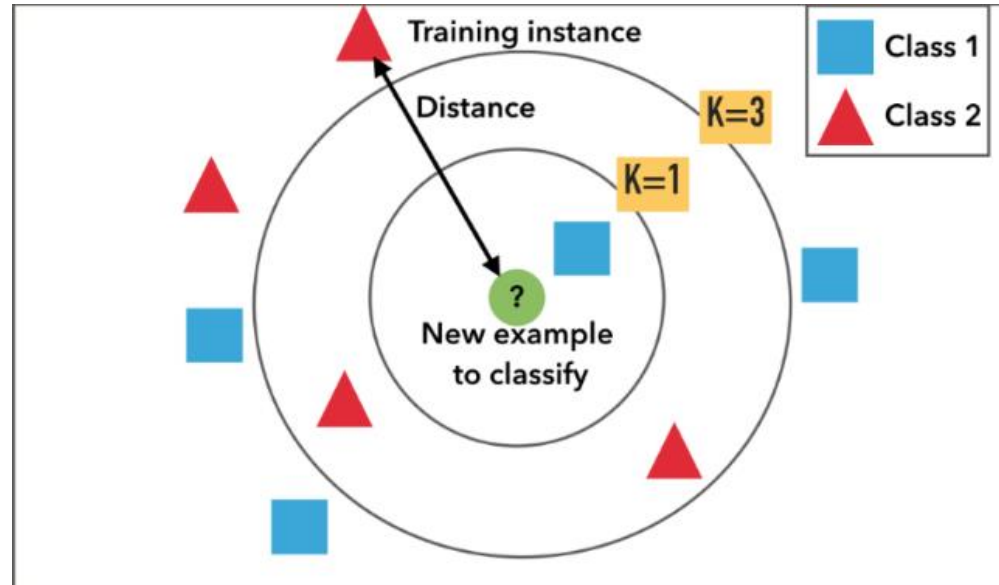
Örüntü Tanıma

Dr. Öğr. Üyesi Mehmet Zahid YILDIRIM

e-mail: m.zahidyildirim@karabuk.edu.tr

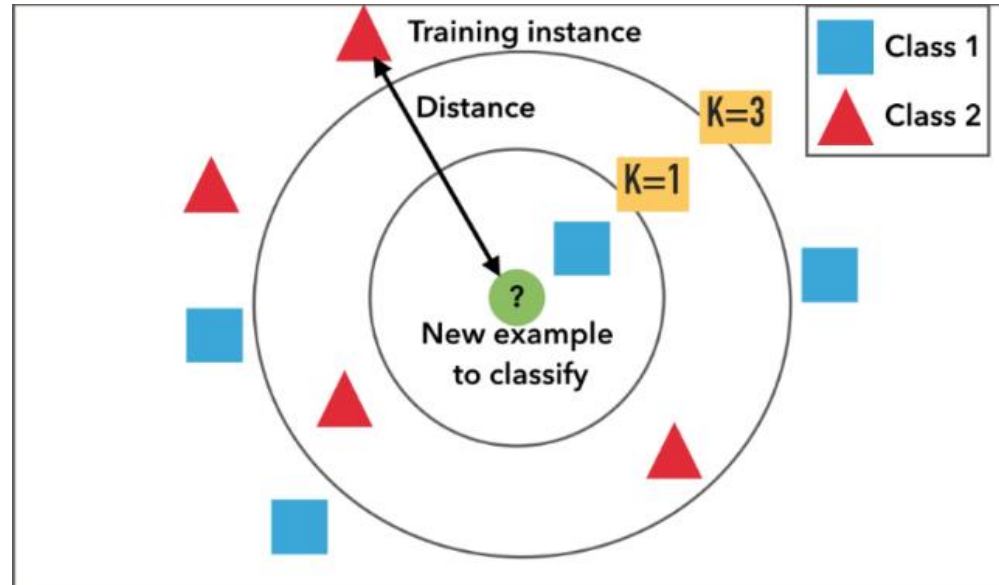
K-En Yakın Komşu (KNN)

- KNN(K-Nearest neighbours algorithm) algoritması veri madenciliği, örüntü tanıma, görüntü işleme gibi bir çok alanda uygulama sahası bulunan bir **sınıflandırma** algoritmasıdır.
- KNN algoritması elimizde var olan gözlem değerlerine ve ait oldukları sınıfa göre sisteme girilmek istenen verinin hangi grubun üyesi olduğunu bulmaya yönelik çalışır.



K-En Yakın Komşu (KNN)

- Bunun için girilen değerin mevcut örneklere uzaklığı hesaplanır.
- Örnek değerin uzaklıkları küçükten büyüğe sıralanır. Daha sonra K değerine göre örnek değerin hangi guruba ait olacağı saptanır.
- Burada **K değeri** Örnek değere en yakın kaç verinin dikkate alınacağı ile ilgili bir sayıdır.



K-En Yakın Komşu (KNN)

- Amazon veya Netflix gibi E-ticaret ve eğlence şirketlerinin çoğu, satın alınacak ürünleri veya izlenecek filmleri/şovları önerirken KNN'yi kullanır.
- Bu şirketler web sitelerinde daha önce satın aldığınız ürünler veya izlediğiniz filmler gibi kullanıcı davranışları hakkında veri toplar ve KNN uygular.
- Şirketler, mevcut müşteri verilerinizi girecek ve bunları benzer ürünleri satın almış veya benzer filmler izlemiş olan diğer müşterilerle karşılaştıracaktır.
- Algoritmanın bu veri noktasını nasıl sınıflandırdığına bağlı olarak ürünler ve filmler size önerilecektir.

K-En Yakın Komşu (KNN)

KNN'nin Avantajları

- Hızlı hesaplama
- Basit algoritma – yorumlamak için
- Çok yönlü – sınıflandırma ve regresyon için kullanışlıdır
- Yüksek doğruluk
- Veriler hakkında varsayım yok - ek varsayımlar yapmaya veya bir model oluşturmaya gerek yok.

K-En Yakın Komşu (KNN)

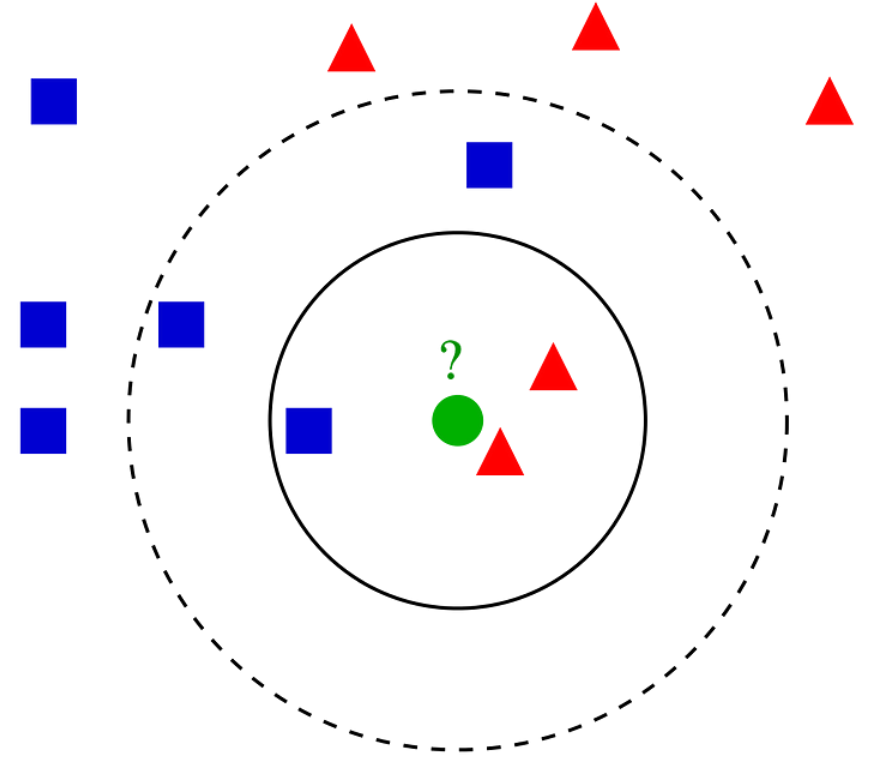
KNN'nin Dezavantajları

- Doğruluk, verilerin kalitesine bağlıdır
- Tahmin, büyük verilerle yavaşlar
- Büyük veri kümeleri için uygun değil
- Tüm eğitim verilerini saklama ihtiyacı, bu nedenle yüksek bellek gerektirir
- Tüm eğitimi sakladığı için hesaplama açısından pahalı olabilir.

K-En Yakın Komşu (KNN)

- KNN algoritmasının adımları şu şekildedir.

1. K parametre değeri belirlenir
2. Sorgu örneği ile tüm eğitim örnekleri arası mesafe hesaplanır.
3. Mesafeler küçükten büyüğe doğru sıralanır.
4. İlk sırada yer alan K tanesi seçilir.
5. Seçilen verilerin kategorileri belirlenir.
6. En çok tekrarlanan kategori verilen örneğin kategorisi olarak belirlenir.



K-En Yakın Komşu (KNN)

- KNN algoritmasında uzaklık hesabı için birçok farklı uzaklık formülü kullanılır.

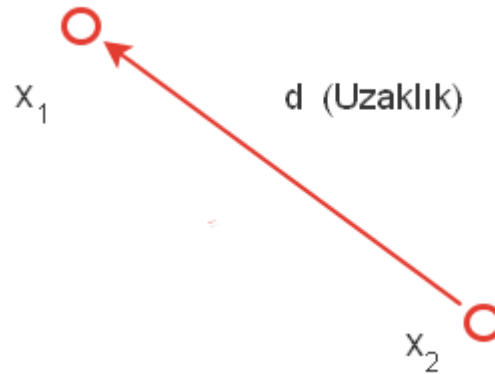
Distance functions

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

K-En Yakın Komşu (KNN)

Euclidean (Öklid) Uzaklık Hesaplaması

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad \longrightarrow \quad d(i, j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + (X_{i3} - X_{j3})^2}$$

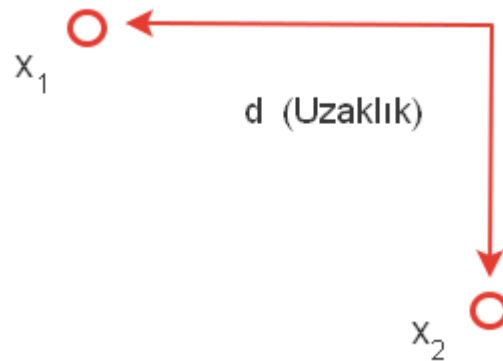


K-En Yakın Komşu (KNN)

Manhattan Uzaklık Hesaplaması

Bu mesafe taksi mesafesi veya blok mesafesi olarak da bilinir, iki nokta arasındaki uzaklık, Kartezyen koordinatlarının mutlak farklarının toplamıdır

$$\sum_{i=1}^k |x_i - y_i| \quad \longrightarrow \quad d(i, j) = |X_{i1} - X_{j1}| + |X_{i2} - X_{j2}| + |X_{i3} - X_{j3}|$$

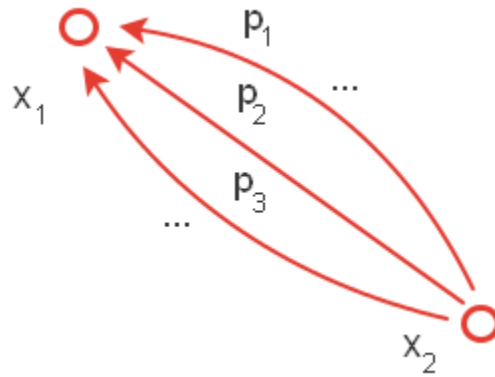


K-En Yakın Komşu (KNN)

Minkowski Uzaklık Hesaplaması

Formülde yer alan 'm' parametresi öznitelik sayısını temsil eder.

$$\left(\sum_{i=1}^k (|x_i - y_i|)^m \right)^{1/m} \longrightarrow d(i, j) = \left[|x_{i1} - x_{j1}|^m + |x_{i2} - x_{j2}|^m + |x_{i3} - x_{j3}|^m \right]^{1/m}$$



K-En Yakın Komşu (KNN)

Gözlem	A	B	C
1	2	3	1
2	4	1	3
3	5	7	3
4	4	8	2
5	3	9	5

Örnek (Öklid):

Burada yer alan üç değişken için, i ve j gözlem noktaları olmak üzere Öklit uzaklık bağıntısını kullanarak 1. ve 2. gözlem arasındaki uzaklık şu şekilde hesaplanır;

$$\begin{aligned}d(2,1) &= \sqrt{(x_{21} - x_{11})^2 + (x_{22} - x_{12})^2 + (x_{23} - x_{13})^2} \\ &= \sqrt{(4 - 2)^2 + (1 - 3)^2 + (3 - 1)^2} = 3.46\end{aligned}$$

3. ve 1. gözlem arasındaki uzaklık;

$$\begin{aligned}d(3,1) &= \sqrt{(x_{31} - x_{11})^2 + (x_{32} - x_{12})^2 + (x_{33} - x_{13})^2} \\ &= \sqrt{(5 - 2)^2 + (7 - 3)^2 + (3 - 1)^2} = 5.39\end{aligned}$$

K-En Yakın Komşu (KNN)

Örnek (Öklid):

Her bir gözlem arasındaki Öklit uzaklıkları hesaplandığında aşağıdaki sonuçlar elde edilir;

Gözlem	A	B	C
1	2	3	1
2	4	1	3
3	5	7	3
4	4	8	2
5	3	9	5

Gözlem	1	2	3	4	5
1	0.00				
2	3.46	0.00			
3	5.39	6.08	0.00		
4	5.48	7.07	1.73	0.00	
5	7.28	8.31	3.46	3.32	0.00

K-En Yakın Komşu (KNN)

Gözlem	A	B	C
1	2	3	1
2	4	1	3
3	5	7	3
4	4	8	2
5	3	9	5

Örnek (Manhattan):

Örnekteki veriler kullanılarak 3 değişken için 2. ve 1. gözlem arasındaki Manhattan uzaklığı aşağıdaki gibi hesaplanır;

$$d(2,1) = |4 - 2| + |1 - 3| + |3 - 1| = 6$$

3. ve 1. gözlem arasındaki uzaklık;

$$d(3,1) = |5 - 2| + |7 - 3| + |3 - 1| = 9$$

K-En Yakın Komşu (KNN)

Örnek (Manhattan):

Her bir gözlem arasındaki Manhattan uzaklıkları hesaplandığında aşağıdaki sonuçlar elde edilir;

Gözlem	A	B	C
1	2	3	1
2	4	1	3
3	5	7	3
4	4	8	2
5	3	9	5

Gözlem	1	2	3	4	5
1	0.00				
2	6.00	0.00			
3	9.00	7.00	0.00		
4	8.00	8.00	3.00	0.00	
5	11.00	11.00	6.00	5.00	0.00

K-En Yakın Komşu (KNN)

Gözlem	A	B	C
1	2	3	1
2	4	1	3
3	5	7	3
4	4	8	2
5	3	9	5

Örnek (Minkowski):

Örnekteki veriler kullanılarak 3 değişken için 2. ve 1. gözlem arasındaki Minkowski uzaklığı aşağıdaki gibi hesaplanır;

$$\begin{aligned}d(2,1) &= \left[|x_{21} - x_{11}|^3 + |x_{22} - x_{12}|^3 + |x_{23} - x_{13}|^3 \right]^{\frac{1}{3}} \\ &= \left[|4 - 2|^3 + |1 - 3|^3 + |3 - 1|^3 \right]^{\frac{1}{3}} = 2.88\end{aligned}$$

3. ve 1. gözlem arasındaki uzaklık;

$$\begin{aligned}d(3,1) &= \left[|x_{31} - x_{11}|^3 + |x_{32} - x_{12}|^3 + |x_{33} - x_{13}|^3 \right]^{\frac{1}{3}} \\ &= \left[|5 - 2|^3 + |7 - 3|^3 + |3 - 1|^3 \right]^{\frac{1}{3}} = 4.63\end{aligned}$$

K-En Yakın Komşu (KNN)

Örnek (Minkowski):

Her bir gözlem arasındaki Minkowski uzaklıkları hesaplandığında aşağıdaki sonuçlar elde edilir;

Gözlem	A	B	C
1	2	3	1
2	4	1	3
3	5	7	3
4	4	8	2
5	3	9	5

Gözlem	1	2	3	4	5
1	0.00				
2	2.88	0.00			
3	4.63	6.01	0.00		
4	5.12	7.01	1.44	0.00	
5	6.55	8.05	2.88	3.07	0.00

K-En Yakın Komşu (KNN)

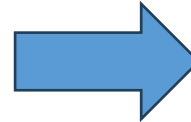
Örnek: Verilen gözlem tablosuna göre yeni bir gözlem olan ($X_1=0.10$, $X_2=0.50$) gözleminin **k=3** için hangi sınıfa dahil olduğunu KNN algoritması ile bulalım.

X_1	X_2	Cinsiyet
0.08	0.20	ERKEK
0.07	0.07	ERKEK
0.20	0.09	ERKEK
1.00	0.20	KADIN
0.05	0.06	ERKEK
0.20	0.25	ERKEK
0.17	0.07	ERKEK
0.15	0.55	KADIN
0.50	0.08	ERKEK
0.10	0.06	KADIN

K-En Yakın Komşu (KNN)

Örnek: Verilen gözlem tablosuna göre yeni bir gözlem olan ($X_1=0.10$, $X_2=0.50$) gözleminin **k=3** için hangi sınıfa dahil olduğunu KNN algoritması ile bulalım.

X_1	X_2	Cinsiyet
0.08	0.20	ERKEK
0.07	0.07	ERKEK
0.20	0.09	ERKEK
1.00	0.20	KADIN
0.05	0.06	ERKEK
0.20	0.25	ERKEK
0.17	0.07	ERKEK
0.15	0.55	KADIN
0.50	0.08	ERKEK
0.10	0.06	KADIN

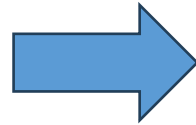


X_1	X_2	Uzaklık
0.08	0.20	0.30
0.07	0.07	0.43
0.20	0.09	0.42
1.00	0.20	0.95
0.05	0.06	0.44
0.20	0.25	0.27
0.17	0.07	0.43
0.15	0.55	0.07
0.50	0.08	0.58
0.10	0.06	0.44

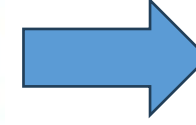
K-En Yakın Komşu (KNN)

Örnek: Verilen gözlem tablosuna göre yeni bir gözlem olan ($X_1=0.10$, $X_2=0.50$) gözleminin $k=3$ için hangi sınıfa dahil olduğunu KNN algoritması ile bulalım.

X_1	X_2	Cinsiyet
0.08	0.20	ERKEK
0.07	0.07	ERKEK
0.20	0.09	ERKEK
1.00	0.20	KADIN
0.05	0.06	ERKEK
0.20	0.25	ERKEK
0.17	0.07	ERKEK
0.15	0.55	KADIN
0.50	0.08	ERKEK
0.10	0.06	KADIN



X_1	X_2	Uzaklık
0.08	0.20	0.30
0.07	0.07	0.43
0.20	0.09	0.42
1.00	0.20	0.95
0.05	0.06	0.44
0.20	0.25	0.27
0.17	0.07	0.43
0.15	0.55	0.07
0.50	0.08	0.58
0.10	0.06	0.44

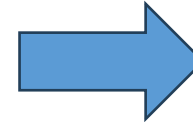


X_1	X_2	Uzaklık	Sıra
0.08	0.20	0.30	3
0.07	0.07	0.43	5
0.20	0.09	0.42	4
1.00	0.20	0.95	10
0.05	0.06	0.44	7
0.20	0.25	0.27	2
0.17	0.07	0.43	6
0.15	0.55	0.07	1
0.50	0.08	0.58	9
0.10	0.06	0.44	8

K-En Yakın Komşu (KNN)

Örnek: Verilen gözlem tablosuna göre yeni bir gözlem olan ($X_1=0.10$, $X_2=0.50$) gözleminin **k=3** için hangi sınıfa dahil olduğunu KNN algoritması ile bulalım.

X_1	X_2	Uzaklık	Sıra	Cinsiyet
0.08	0.20	0.30	3	ERKEK
0.07	0.07	0.43	5	
0.20	0.09	0.42	4	
1.00	0.20	0.95	10	
0.05	0.06	0.44	7	
0.20	0.25	0.27	2	ERKEK
0.17	0.07	0.43	6	
0.15	0.55	0.07	1	KADIN
0.50	0.08	0.58	9	
0.10	0.06	0.44	8	



Sınıf → ERKEK

K-En Yakın Komşu (KNN)

Örnek: Verilen gözlem tablosuna göre yeni bir gözlem olan ($X_1=8$, $X_2=4$) gözleminin **k=4** için hangi sınıfa dahil olduğunu KNN algoritması ile bulalım.

X_1	X_2	Y
2	4	KÖTÜ
3	6	İYİ
3	4	İYİ
4	10	KÖTÜ
5	8	KÖTÜ
6	3	İYİ
7	9	İYİ
9	7	KÖTÜ
11	7	KÖTÜ
10	2	KÖTÜ

K-En Yakın Komşu (KNN)

Örnek:

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

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

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

$$d(i, j) = \sqrt{(4-8)^2 + (10-4)^2} = 7.21$$

$$d(i, j) = \sqrt{(5-8)^2 + (8-4)^2} = 5.00$$

$$d(i, j) = \sqrt{(6-8)^2 + (3-4)^2} = 2.24$$

$$d(i, j) = \sqrt{(7-8)^2 + (9-4)^2} = 5.10$$

$$d(i, j) = \sqrt{(9-8)^2 + (7-4)^2} = 3.16$$

$$d(i, j) = \sqrt{(11-8)^2 + (7-4)^2} = 4.24$$

$$d(i, j) = \sqrt{(10-8)^2 + (2-4)^2} = 2.83$$

X1	X2	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

K-En Yakın Komşu (KNN)

Örnek:

X1	X2	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

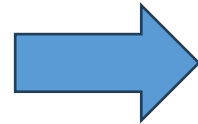


X1	X2	Uzaklık	Sıra
2	4	6.00	9
3	6	5.39	8
3	4	5.00	6
4	10	7.21	10
5	8	5.00	5
6	3	2.24	1
7	9	5.10	7
9	7	3.16	3
11	7	4.24	4
10	2	2.83	2

K-En Yakın Komşu (KNN)

Örnek:

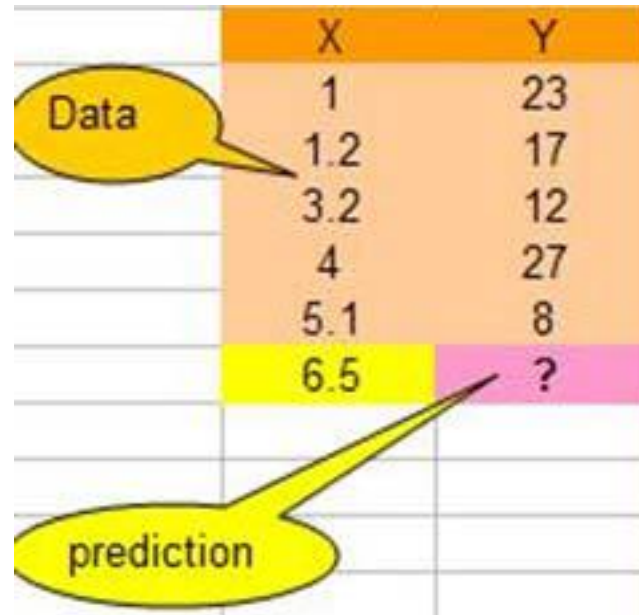
X1	X2	Uzaklık	Sıra	Y
2	4	6.00	9	
3	6	5.39	8	
3	4	5.00	6	
4	10	7.21	10	
5	8	5.00	5	
6	3	2.24	1	İYİ
7	9	5.10	7	
9	7	3.16	3	KÖTÜ
11	7	4.24	4	KÖTÜ
10	2	2.83	2	KÖTÜ



Sınıf → KÖTÜ

K-En Yakın Komşu (KNN)

KNN Algoritmasının Tahmin için Kullanımı

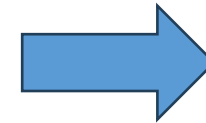


	X	Y
Data	1	23
	1.2	17
	3.2	12
	4	27
	5.1	8
	6.5	?
prediction		

K-En Yakın Komşu (KNN)

KNN Algoritmasının Tahmin için Kullanımı

X	Y	distance	Nearest Neighbor Value
1	23	5.5	27 8
1.2	17	5.3	
3.2	12	3.3	
4	27	2.5	
5.1	8	1.4	



$$\frac{27+8}{2}=17.5$$

K-En Yakın Komşu (KNN)

KNN Algoritmasının Tahmin için Kullanımı

K-Nearest Neighbor for Time Series				
	K	2		
	X	Y	distance	Nearest Neighbor Value
Data	1	23	5.5	27 8
	1.2	17	5.3	
	3.2	12	3.3	
	4	27	2.5	
	5.1	8	1.4	
	6.5	?		
prediction				
result				
KNN prediction				17.5