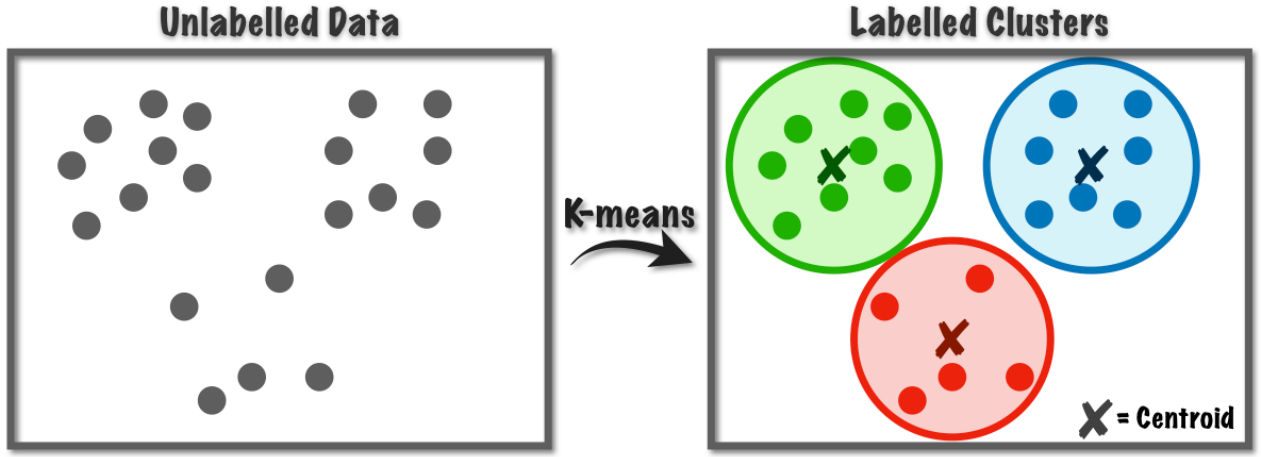
Örüntü Tanıma

Dr. Öğr. Üyesi Mehmet Zahid YILDIRIM

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

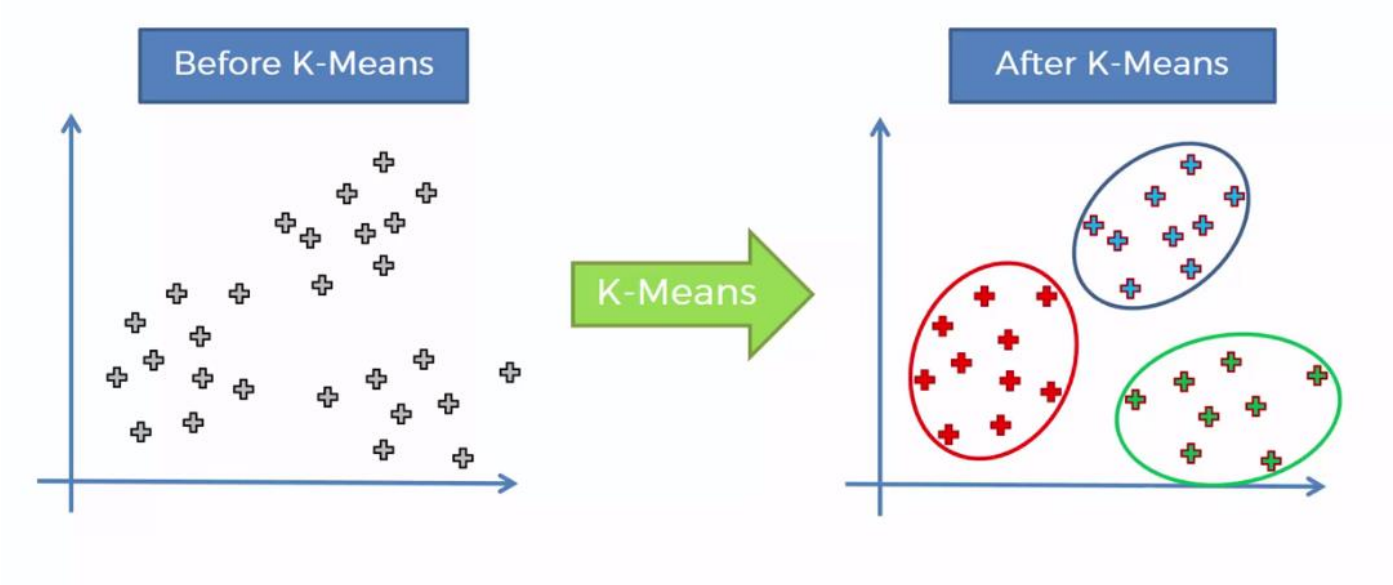
K-Ortalama (K-Means)

- K-Means algoritması bir unsupervised learning (gözetimsiz öğrenme) ve **kümeleme** algoritmasıdır.
- Yöntem N adet veri nesnesinden oluşan bir veri kümesini giriş parametresi olarak verilen **K** adet kümeye bölümler.
- Amaç, gerçekleştirilen bölümleme işlemi sonunda elde edilen kümelerin, küme içi benzerliklerinin **maksimum** ve kümeler arası benzerliklerinin ise **minimum** olmasını sağlamaktır.



K-Ortalama (K-Means)

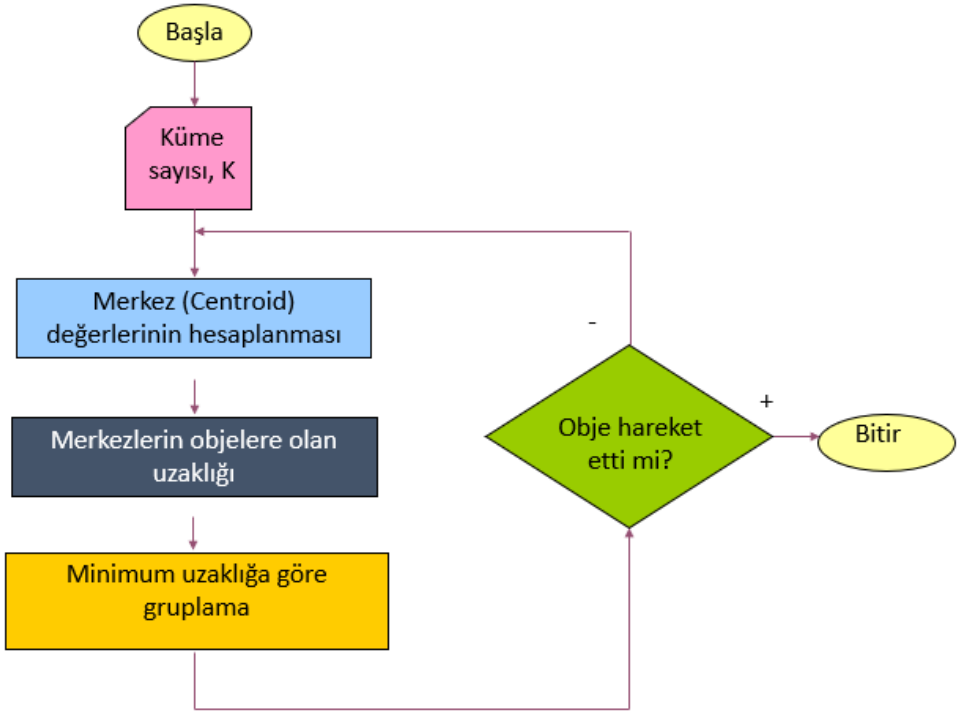
- Farklı bir ifadeyle K-Means, eldeki verileri özelliklerine göre hiçbir sınıf bilgisi olmadan K sayıda kümeye gruplama işlemidir.



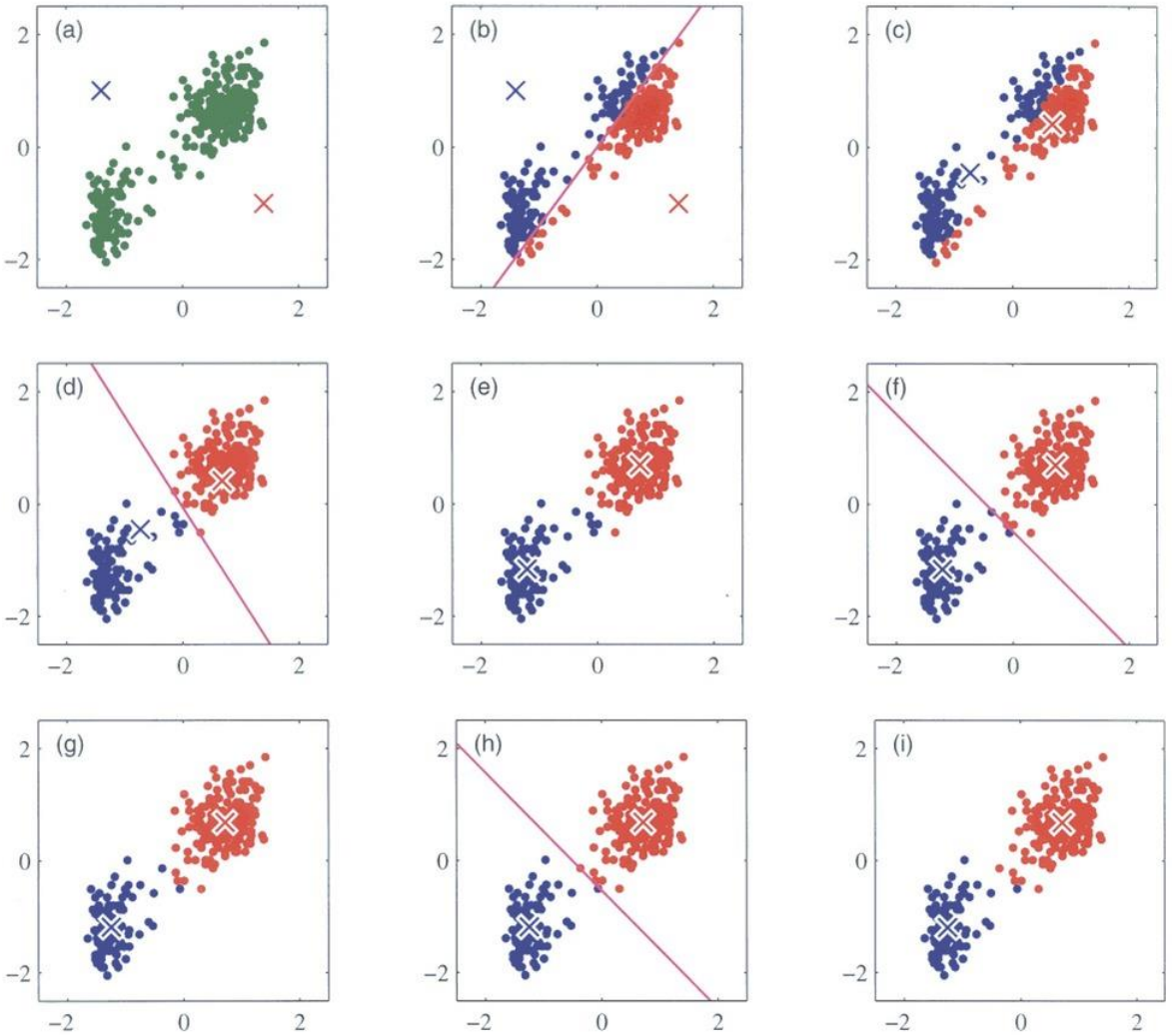
K-Ortalama (K-Means)

- K-Means algoritmasının adımları şu şekildedir.

1. K parametre değeri belirlenir
2. K adet rasgele merkez seçilir.
3. Seçilen veriler hangi merkeze yakınsa o gruba dahil edilir.
4. Ayrılan grupların ortalaması bulunur.
5. 3. ve 4. adımlar merkez noktalar yer değiştirmedikçe tekrar edilir.

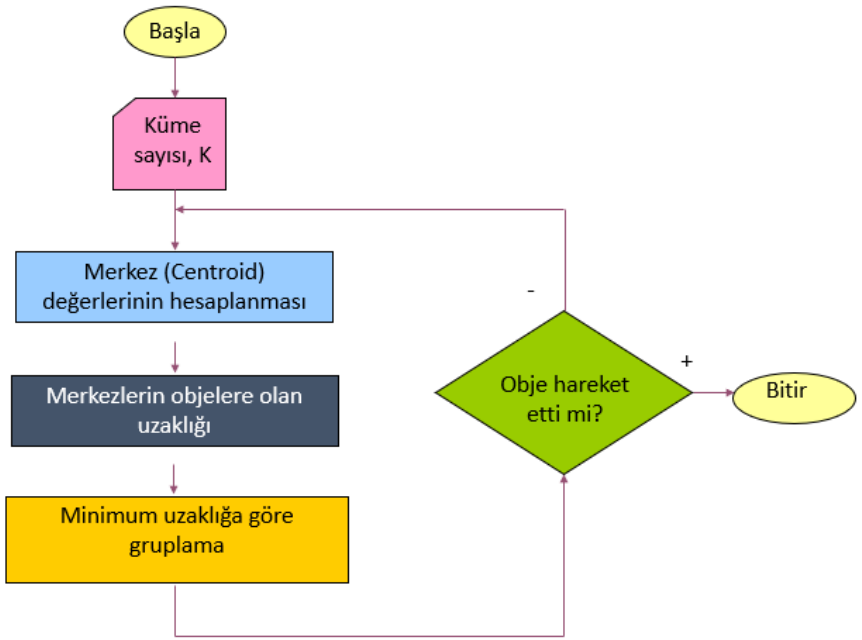


K-Ortalama (K-Means)



K-Ortalama (K-Means)

Örnek: Verilen veri tablosuna göre **K=2** için K-Means algoritması ile kümeleme yapınız.



Instances (Örnekler)	X	Y
1	1.0	1.5
2	1.0	4.5
3	2.0	1.5
4	2.0	3.5
5	3.0	2.5
6	5.0	6.0

K-Ortalama (K-Means)

Adım 1 : Ortalamayı hesaplamak için rastgele iki nokta belirlenir.

Bu noktalar, **C1** : (1.0 , 1.5) ve **C2** : (2.0 , 1.5) olsun.

Instances (Örnekler)	X	Y
1	1.0	1.5
2	1.0	4.5
3	2.0	1.5
4	2.0	3.5
5	3.0	2.5
6	5.0	6.0

K-Ortalama (K-Means)

Adım 2 : Öklit, Manhattan veya Minkowski uzaklık formüllerinden biriyle her bir satır için hesaplama yapılır.

C1

$d(c_1, I1) = \sqrt{(1.0 - 1.0)^2 + (1.5 - 1.5)^2} = 0$

$d(c_1, I2) = \sqrt{(1.0 - 1.0)^2 + (1.5 - 4.5)^2} = 3$

$d(c_1, I3) = \sqrt{(1.0 - 2.0)^2 + (1.5 - 1.5)^2} = 2.24$

$d(c_1, I4) = \sqrt{(1.0 - 2.0)^2 + (1.5 - 3.5)^2} = 2.24$

$d(c_1, I5) = \sqrt{(1.0 - 3.0)^2 + (1.5 - 2.5)^2} = 2.24$

$d(c_1, I6) = \sqrt{(1.0 - 2.0)^2 + (1.5 - 1.5)^2} = 6.02$

C2

$d(c_2, I1) = \sqrt{(2.0 - 1.0)^2 + (1.5 - 1.5)^2} = 1$

$d(c_2, I2) = \sqrt{(2.0 - 1.0)^2 + (1.5 - 4.5)^2} = 3.1$

$d(c_2, I3) = \sqrt{(2.0 - 2.0)^2 + (1.5 - 1.5)^2} = 0$

$d(c_2, I4) = \sqrt{(2.0 - 2.0)^2 + (1.5 - 3.5)^2} = 2$

$d(c_2, I5) = \sqrt{(2.0 - 3.0)^2 + (1.5 - 2.5)^2} = 1.4$

$d(c_2, I6) = \sqrt{(2.0 - 2.0)^2 + (1.5 - 1.5)^2} = 5.4$

Instances (Örnekler)	X	Y
1	1.0	1.5
2	1.0	4.5
3	2.0	1.5
4	2.0	3.5
5	3.0	2.5
6	5.0	6.0

K-Ortalama (K-Means)

Adım 2 : Öklit, Manhattan veya Minkowski uzaklık formüllerinden biriyle her bir satır için hesaplama yapılır.

C1		C2
$d(c_1, I1) = \sqrt{(1.0-1.0)^2+(1.5-1.5)^2} = 0$	<	$d(c_2, I1) = \sqrt{(2.0-1.0)^2+(1.5-1.5)^2} = 1$
$d(c_1, I2) = \sqrt{(1.0-1.0)^2+(1.5-4.5)^2} = 3$	<	$d(c_2, I2) = \sqrt{(2.0-1.0)^2+(1.5-4.5)^2} = 3.1$
$d(c_1, I3) = \sqrt{(1.0-2.0)^2+(1.5-1.5)^2} = 2.24$	>	$d(c_2, I3) = \sqrt{(2.0-2.0)^2+(1.5-1.5)^2} = 0$
$d(c_1, I4) = \sqrt{(1.0-2.0)^2+(1.5-3.5)^2} = 2.24$	>	$d(c_2, I4) = \sqrt{(2.0-2.0)^2+(1.5-3.5)^2} = 2$
$d(c_1, I5) = \sqrt{(1.0-3.0)^2+(1.5-2.5)^2} = 2.24$	>	$d(c_2, I5) = \sqrt{(2.0-3.0)^2+(1.5-2.5)^2} = 1.4$
$d(c_1, I6) = \sqrt{(1.0-2.0)^2+(1.5-1.5)^2} = 6.02$	>	$d(c_2, I6) = \sqrt{(2.0-2.0)^2+(1.5-1.5)^2} = 5.4$

I1 , I2

I3 , I4 , I5 , I6

Instances (Örnekler)	X	Y
1	1.0	1.5
2	1.0	4.5
3	2.0	1.5
4	2.0	3.5
5	3.0	2.5
6	5.0	6.0

K-Ortalama (K-Means)

Adım 3 : Kümelenen örneklerin nitelikleri ile ortalama hesaplanır.

I1 , I2

$C_1 = (X_1 + X_2 / 2 , Y_1 + Y_2 / 2)$

$C_1 = (1.0 + 1.0 / 2 , 1.5 + 4.5 / 2)$

$C_1 = (1 , 3)$

I3 , I4 , I5 , I6

$C_2 = (X3 + X4 + X5 + X6 / 4 , Y3 + Y4 + Y5 + Y6 / 4)$

$C_2 = (2.0 + 2.0 + 3.0 + 5.0 / 4 , 1.5 + 3.5 + 2.5 + 6.0 / 4)$

$C_2 = (3 , 3.38)$

Instances (Örnekler)	X	Y
1	1.0	1.5
2	1.0	4.5
3	2.0	1.5
4	2.0	3.5
5	3.0	2.5
6	5.0	6.0

K-Ortalama (K-Means)

Adım 3 : Kümelenen örneklerin nitelikleri ile ortalama hesaplanır.

C_1

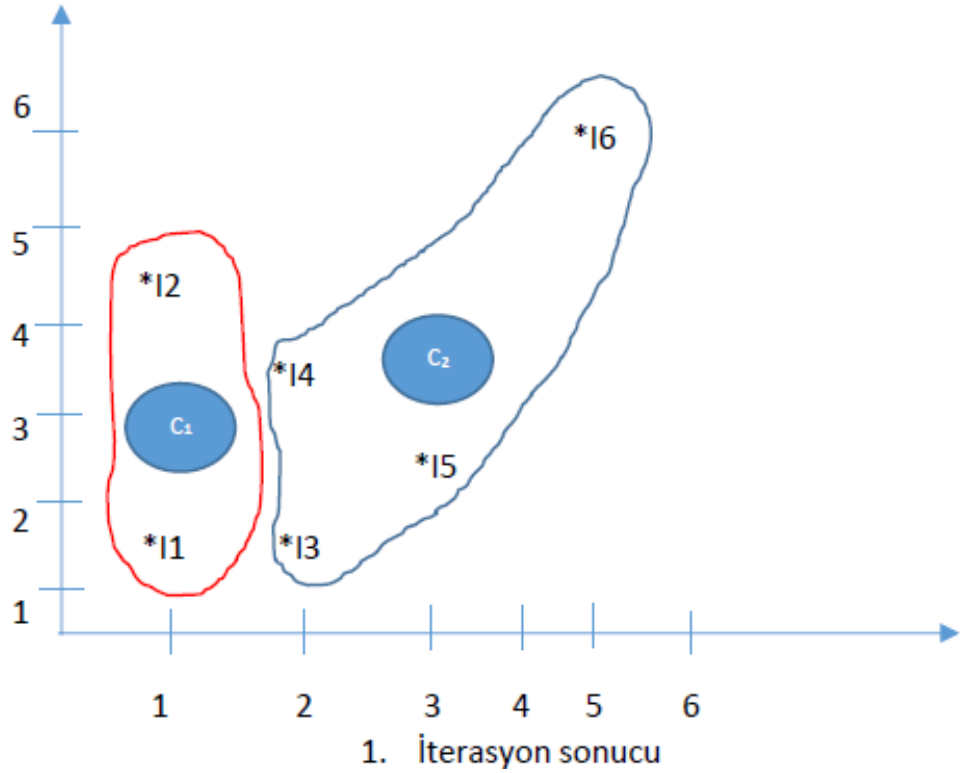
I1 , I2

$C_1 = (X_1 + X_2 / 2 , Y_1 + Y_2 / 2)$
$C_1 = (1.0 + 1.0 / 2 , 1.5 + 4.5 / 2)$
$C_1 = (1 , 3)$

C_2

I3 , I4 , I5 , I6

$C_2 = (X3 + X4 + X5 + X6 / 4 , Y3 + Y4 + Y5 + Y6 / 4)$
$C_2 = (2.0 + 2.0 + 3.0 + 5.0 / 4 , 1.5 + 3.5 + 2.5 + 6.0 / 4)$
$C_2 = (3 , 3.38)$



1. İterasyon sonucu

K-Ortalama (K-Means)

Adım 4 : İlk iterasyon tamamlandıktan sonra 2. İterasyon için yeni C1 ve C2 noktaları kullanılarak tekrar kümeleme işlemi yapılır ve ortalamalar bulunur..

$d(c_1, I1) = \sqrt{(1.0-1.0)^2+(3.0-1.5)^2} = 1.5 < d(c_2, I1) = \sqrt{(3.0-1.0)^2+(3.38-1.5)^2} = 2.75$

$d(c_1, I2) = \sqrt{(1.0-1.0)^2+(3.0-4.5)^2} = 1.5 < d(c_2, I2) = \sqrt{(3.0-1.0)^2+(3.38-4.5)^2} = 2.25$

$d(c_1, I3) = \sqrt{(1.0-2.0)^2+(3.0-1.5)^2} = 1.8 < d(c_2, I3) = \sqrt{(3.0-2.0)^2+(3.38-1.5)^2} = 2.14$

$d(c_1, I4) = \sqrt{(1.0-2.0)^2+(3.0-3.5)^2} = 1.11 > d(c_2, I4) = \sqrt{(3.0-2.0)^2+(3.38-3.5)^2} = 1.0$

$d(c_1, I5) = \sqrt{(1.0-3.0)^2+(3.0-2.5)^2} = 2.06 > d(c_2, I5) = \sqrt{(3.0-3.0)^2+(3.38-2.5)^2} = 0.90$

$d(c_1, I6) = \sqrt{(1.0-2.0)^2+(3.0-1.5)^2} = 5.0 > d(c_2, I6) = \sqrt{(3.0-2.0)^2+(3.38-1.5)^2} = 3.20$

I1 , I2 , I3

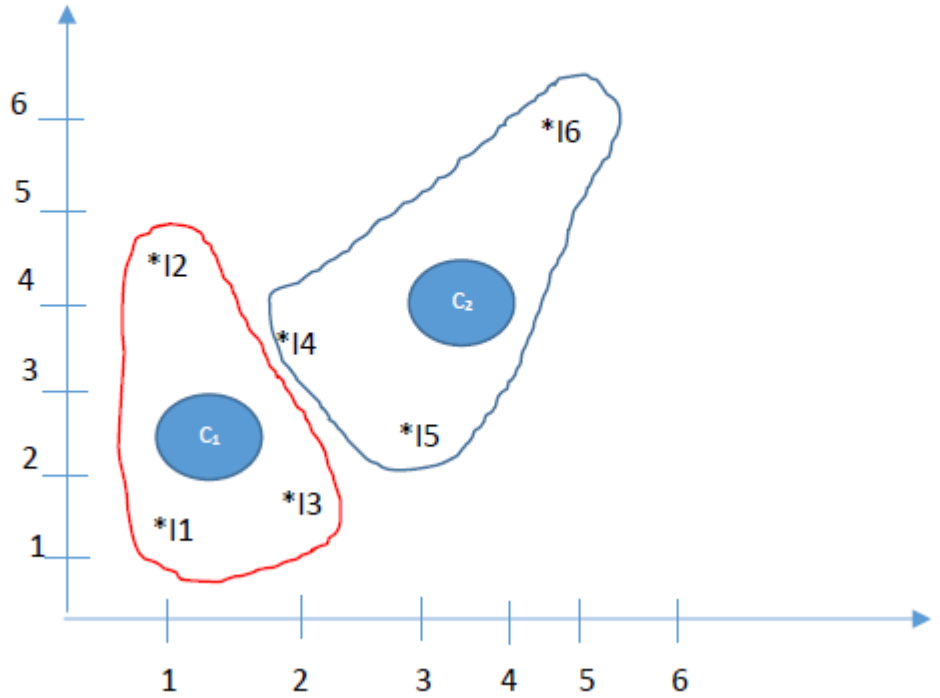
I4 , I5 , I6

Instances (Örnekler)	X	Y
1	1.0	1.5
2	1.0	4.5
3	2.0	1.5
4	2.0	3.5
5	3.0	2.5
6	5.0	6.0

K-Ortalama (K-Means)

$C_1 = (X_1 + X_2 + X_3 / 3 , Y_1 + Y_2 + Y_3 / 3)$
$C_1 = (1.0 + 1.0 + 2.0 / 3, 1.5 + 4.5 + 1.5 / 3)$
$C_1 = (1.33 , 2.5)$

$C_2 = (X4 + X5 + X6 / 3 , Y4 + Y5 + Y6 / 3)$
$C_2 = (2.0 + 3.0 + 5.0 / 3 , 3.5 + 2.5 + 6.0 / 3)$
$C_2 = (3.33 , 4)$



Instances (Örnekler)	X	Y
1	1.0	1.5
2	1.0	4.5
3	2.0	1.5
4	2.0	3.5
5	3.0	2.5
6	5.0	6.0

$d(c_1, I1) = \sqrt{(1.0-1.0)^2+(3.0-1.5)^2} = 1.5 < d(c_2, I1) = \sqrt{(3.0-1.0)^2+(3.38-1.5)^2} = 2.75$

$d(c_1, I2) = \sqrt{(1.0-1.0)^2+(3.0-4.5)^2} = 1.5 < d(c_2, I2) = \sqrt{(3.0-1.0)^2+(3.38-4.5)^2} = 2.25$

$d(c_1, I3) = \sqrt{(1.0-2.0)^2+(3.0-1.5)^2} = 1.8 < d(c_2, I3) = \sqrt{(3.0-2.0)^2+(3.38-1.5)^2} = 2.14$

$d(c_1, I4) = \sqrt{(1.0-2.0)^2+(3.0-3.5)^2} = 1.11 > d(c_2, I4) = \sqrt{(3.0-2.0)^2+(3.38-3.5)^2} = 1.0$

$d(c_1, I5) = \sqrt{(1.0-3.0)^2+(3.0-2.5)^2} = 2.06 > d(c_2, I5) = \sqrt{(3.0-3.0)^2+(3.38-2.5)^2} = 0.90$

$d(c_1, I6) = \sqrt{(1.0-2.0)^2+(3.0-1.5)^2} = 5.0 > d(c_2, I6) = \sqrt{(3.0-2.0)^2+(3.38-1.5)^2} = 3.20$

C_1

I1 , I2 , I3

C_2

I4 , I5 , I6

$C_1 = (X_1 + X_2 + X_3 / 3 , Y_1 + Y_2 + Y_3 / 3)$

$C_1 = (1.0 + 1.0 + 2.0 / 3, 1.5 + 4.5 + 1.5 / 3)$

$C_1 = (1.33 , 2.5)$

$C_2 = (X4 + X5 + X6 / 3 , Y4 + Y5 + Y6 / 3)$

$C_2 = (2.0 + 3.0 + 5.0 / 3 , 3.5 + 2.5 + 6.0 / 3)$

$C_2 = (3.33 , 4)$