

# Makine Öğrenmesi

## Güdümsüz Öğrenme

**İlker Birbil ve Utku Karaca**

Erasmus Üniversitesi Rotterdam

İstanbul’da Makine Öğrenmesi

27 Ocak – 2 Şubat, 2020



# Makine Öğrenmesi

```
graph TD; A[Makine Öğrenmesi] --> B[Doğrusal Bağlanım]; A --> C[Boyut Küçültme ve Düzenleştirme]; A --> D[Tekrar Örneklem ve Model Değerlendirme]; A --> E[Sınıflandırma ve Ağaçlar]; A --> F[GÜdÜmsüz Öğrenme]; A --> G[Yapay Sinir Ağları ve Derin Öğrenme]; A --> H[ ];
```

Doğrusal Bağlanım

Boyut Küçültme  
ve  
Düzenleştirme

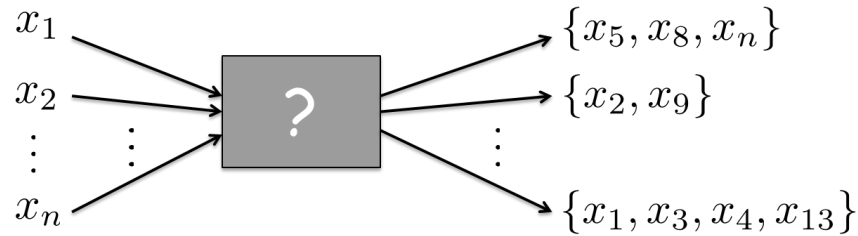
Tekrar Örneklem  
ve  
Model Değerlendirme

Sınıflandırma  
ve  
Ağaçlar

GÜdÜmsüz  
Öğrenme

Yapay Sinir Ağları  
ve  
Derin Öğrenme

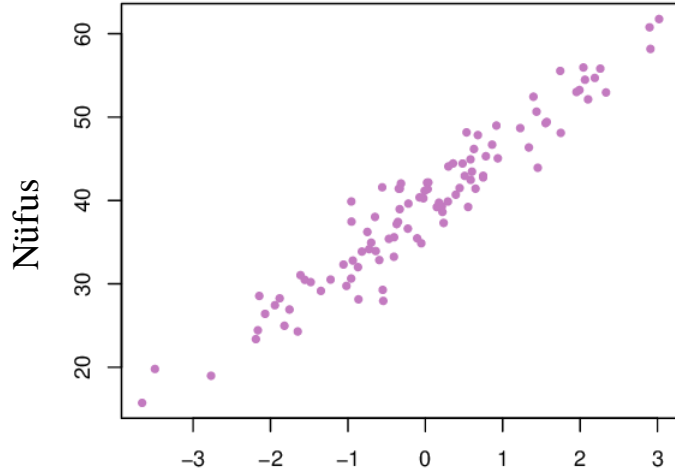
# Güdümsüz Öğrenme



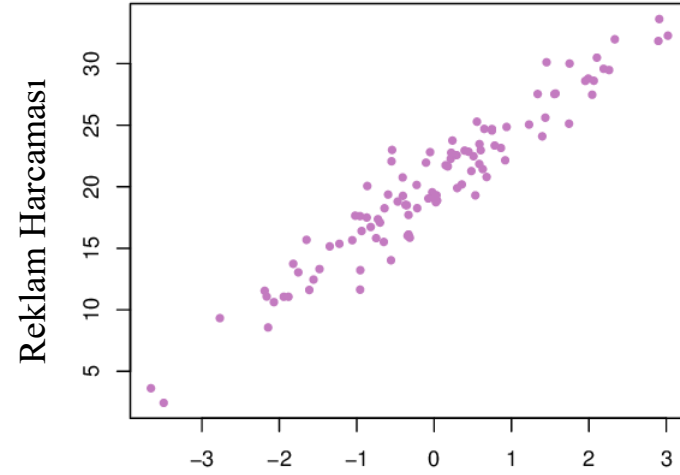
eğitim verisi  
 $\{x_i : 1, \dots, n\}$

- Ana Bileşenler Analizi (Principal Component Analysis): Varyansın büyük kısmını göz önüne alacak şekilde problemin boyutunu düşürme
- Gruplandırma (clustering): Veri kümesi içerisinde benzer alt grupları bulmak
  - K-ortalamalar (K-means) gruplandırma
  - Hiyerarşik (hierarchical) gruplandırma
  - Spektral (spectral) gruplandırma
- Matris Ayırıştırma
- Google PageRank Algoritması

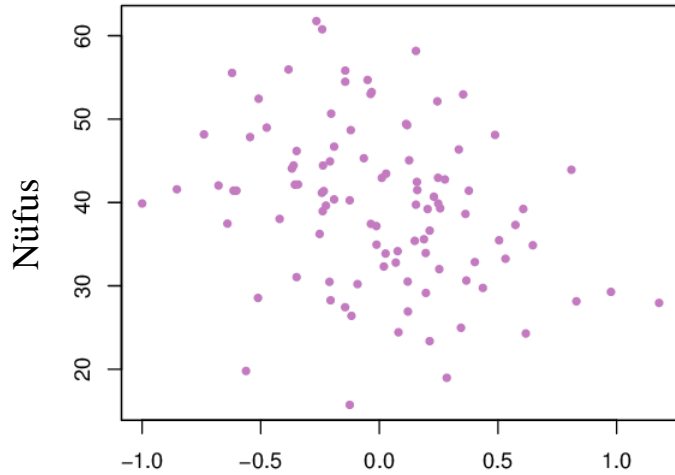
# Resimlerle Ana Bileşenler Analizi



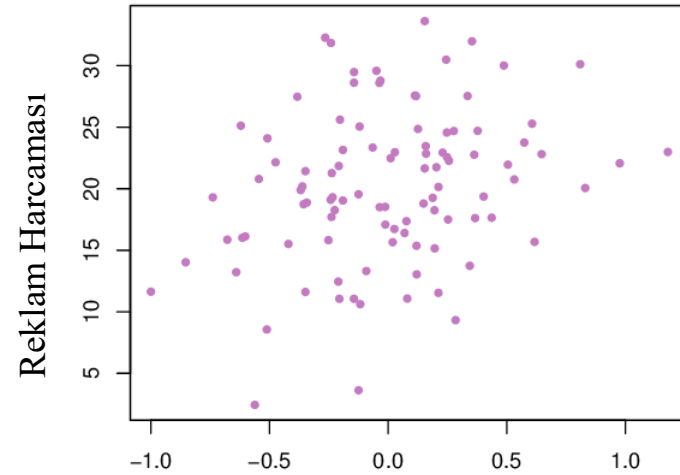
1. Ana Bileşen



1. Ana Bileşen

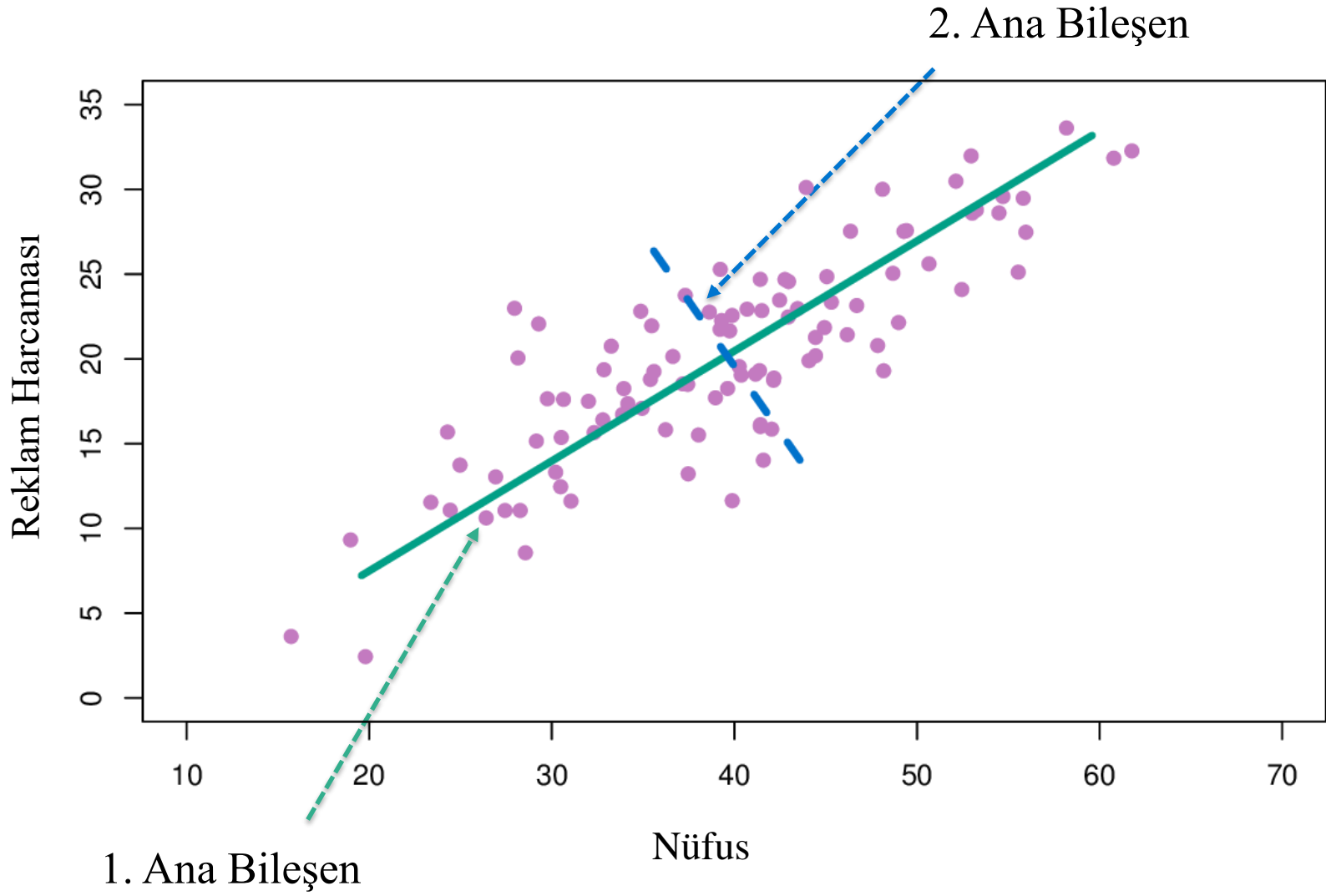


2. Ana Bileşen



2. Ana Bileşen

# Resimlerle Ana Bileşenler Analizi



# Ana Bileşenler Analizi

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \cdots + \phi_{p1}X_p$$

Değişkenlerin ortalamalarının sıfır olacağı kabul edilir

birinci ana  
bileşenin  
skorları

$$z_{i1} = \phi_{11}x_{i1} + \phi_{21}x_{i2} + \cdots + \phi_{p1}x_{ip}$$

Birinci ana bileşenin *ağırlıkları* (loadings)

Varyansı en fazla yapacak doğrultuyu bulma

$$\text{enbüyük} \quad \frac{1}{n} \sum_{i=1}^n \underbrace{(\phi_1^\top x_i)^2}_{z_{i1}^2}$$

$\phi_1 \in \mathbb{R}^p$

$$\text{öyle ki} \quad \|\phi_1\|^2 = 1.$$

# Ana Bileşenler Analizi

$$X^T = [x_1, x_2, \dots, x_n]$$

enbüyükle $\phi_1 \in \mathbb{R}^p$ öyle ki	$\frac{1}{n} \sum_{i=1}^n \underbrace{(\phi_1^T x_i)^2}_{z_{i1}^2}$ $\ \phi_1\ ^2 = 1.$	$\equiv$	enbüyükle $\phi_1 \in \mathbb{R}^p$ öyle ki	$\frac{1}{n} \phi_1^T X^T X \phi_1$ $\phi_1^T \phi_1 = 1.$
---	--	----------	---	---

enbüyükle  
 $\phi_1 \in \mathbb{R}^p$   
öyle ki

$$\frac{1}{n} \phi_1^T X^T X \phi_1$$

$$\phi_1^T \phi_1 = 1.$$

$\phi_1$

enbüyükle  
 $\phi_2 \in \mathbb{R}^p$   
öyle ki

$$\frac{1}{n} \phi_2^T X^T X \phi_2$$

$$\phi_2^T \phi_2 = 1,$$

$$\phi_1^T \phi_2 = 0.$$

Birinci ana bileşenle  
korelasyona engel  
olmak için eklenir  
(diklik kısıtı)

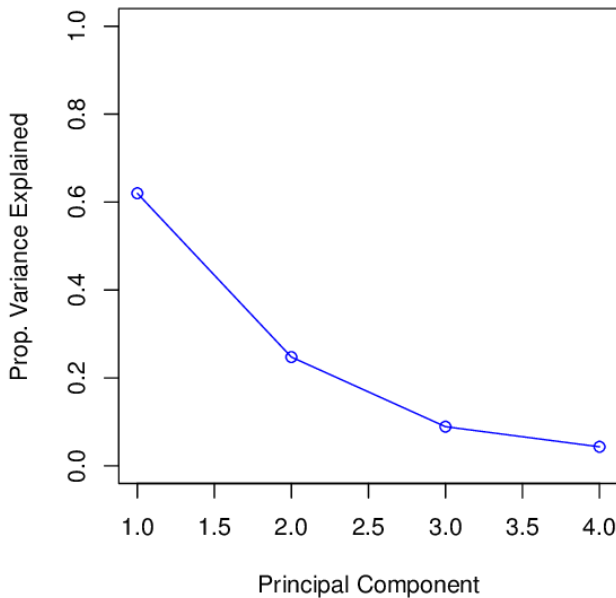
Ana bileşen doğrultuları  $(\phi_1, \phi_2, \dots)$ , buradaki  $X^T X$  matrisinin sıralı özvektörlerine karşılık gelir.

# Ana Bileşenler Analizi

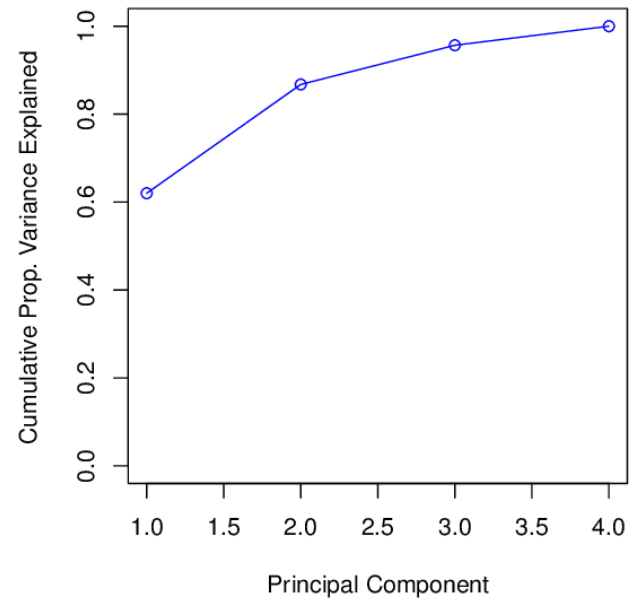
## Varyansın Açıklanması

$$\sum_{i=1}^p \text{Var}(X_j) = \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n x_{ij}^2$$
$$\frac{1}{n} \sum_{i=1}^n z_{im}^2 = \frac{1}{n} \sum_{i=1}^n \left( \sum_{j=1}^p \phi_{jm} x_{ij} \right)^2$$

$$\frac{\sum_{i=1}^n \left( \sum_{j=1}^p \phi_{jm} x_{ij} \right)^2}{\sum_{j=1}^p \sum_{i=1}^n x_{ij}^2}$$



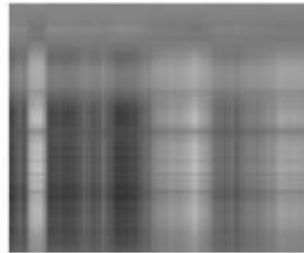
“kayışat (scree) grafiği”





# Ana Bileşenler Analizi

## Görüntü Sıkıştırma



(a) 1 principal component



(b) 5 principal component



(c) 9 principal component



(d) 13 principal component



(e) 17 principal component



(f) 21 principal component



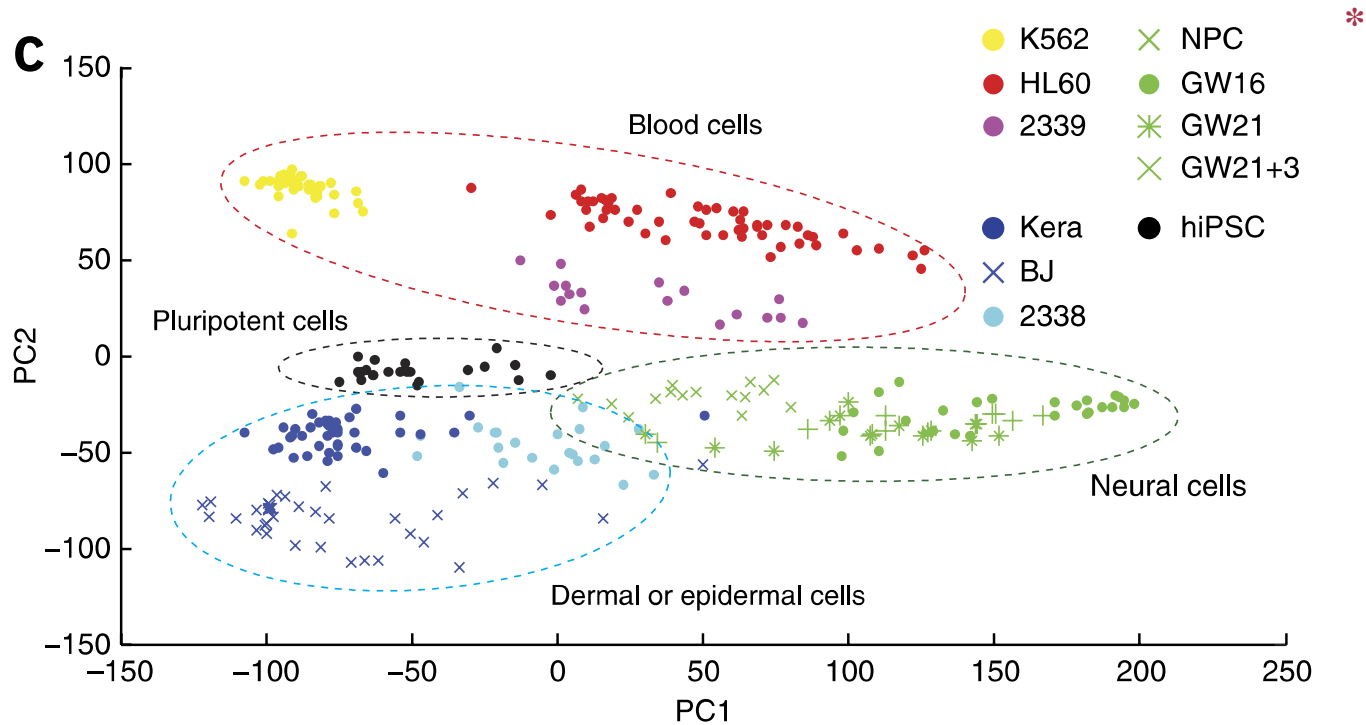
(g) 25 principal component



(h) 29 principal component

# Ana Bileşenler Analizi

## Bioteknoloji



\*

*\*Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex, A. A. Pollen et al., Nature Biotechnology, 32, 1053-1058, 2014.*

# Benzemezlik Ölçüleri

- Kritik Nokta: İki nokta arasındaki uzaklık veya benzemelik
- İyi bir ölçü (genellikle) algoritmadan daha önemlidir
- Bilinmeyen değerler için dikkatli olunmalı
- Algoritmaların çoğu  $K$  küme arasında toplam benzemezliği enazlar:

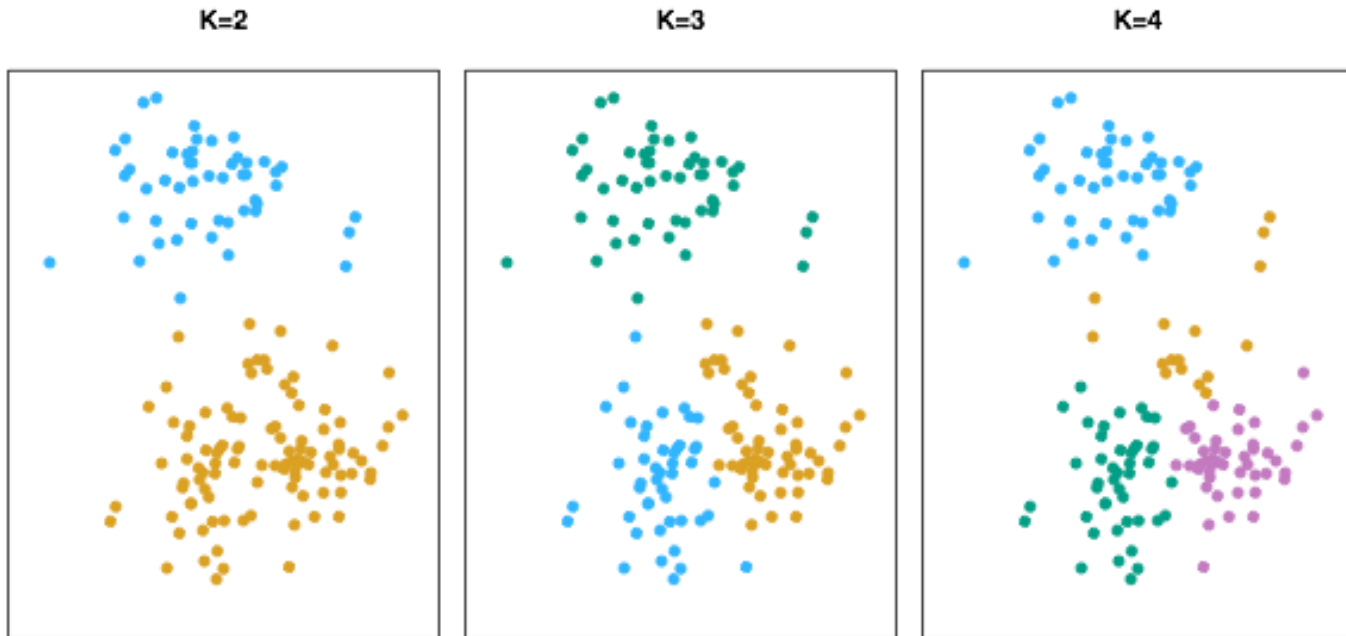
$$W(C) = \frac{1}{2} \sum_{k=1}^K \sum_{C(i)=k} \sum_{C(j)=k} d(x_i, x_j)$$

*kodlayıcı*  
(veri noktalarını kümelere atar)

iki gözlem noktası  
arasındaki benzemezlik

# K-Ortalamalar Gruplandırma

**Fikir:** Birbirine yakın noktaları K gruptan birine atamak



Optimizasyon problemi?

# K-Ortalamalar Gruplandırma

**Performans:** Her noktanın ait olduğu grup merkezine (geometrik merkez) olan mesafesini hesapla ve tüm mesafeleri topla

$$z_{ik} = \begin{cases} 1, & \text{eğer } i \text{ noktası } k \text{ grubunda ise;} \\ 0, & \text{aksi halde.} \end{cases}$$

enküçükle  $\sum_{k=1}^K \sum_{i=1}^n \|x_i - \mu_k\|^2 z_{ik}$   
 $\mu, z$

öyle ki  $\sum_{k=1}^K z_{ik} = 1, \quad i = 1, \dots, n; \quad \text{ZOR!}$   
 $z_{ik} \in \{0, 1\}, \quad i = 1, \dots, n; k = 1, \dots, K;$   
 $\mu_k \in \mathbb{R}^p, \quad k = 1, \dots, K.$

Yaklaşık çözüm?

# K-Ortalamlar Gruplandırma

Her noktayı bir gruba rassal olarak ata

rassal bir  $\bar{k} \in \{1, \dots, K\}$  için  $z_{i\bar{k}}^0 = 1$

$k \in \{1, \dots, K\}$  ve  $k \neq \bar{k}$  için  $z_{ik}^0 = 0$

$t = 1$

Gruplar değiştiği sürece aşağıdaki adımları tekrarla

Çöz

$$\mu^t = \arg \min_{\mu} \sum_{k=1}^K \sum_{i=1}^n \|x_i - \mu_k\|^2 \underbrace{z_{ik}^{t-1}}_{\text{sabit}}$$

Çöz

$$z^t = \arg \min_z \sum_{k=1}^K \sum_{i=1}^n \|x_i - \underbrace{\mu_k^t}_{\text{sabit}}\|^2 z_{ik}$$

$$\text{öyle ki } \sum_{k=1}^K z_{ik} = 1, \quad i = 1, \dots, n;$$

$$z_{ik} \in \{0, 1\}, \quad i = 1, \dots, n; k = 1, \dots, K.$$

$$t \leftarrow t + 1$$

# K-Ortalamalar Gruplandırma

Amaç fonksiyonu her döngüde azalır ya da değerini korur

$$\begin{aligned}\sum_{k=1}^K \sum_{i=1}^n \|x_i - \mu_k^{t-1}\|^2 z_{ik}^{t-1} &\geq \sum_{k=1}^K \sum_{i=1}^n \|x_i - \mu_k^t\|^2 z_{ik}^{t-1} \\ &\geq \sum_{k=1}^K \sum_{i=1}^n \|x_i - \mu_k^t\|^2 z_{ik}^t\end{aligned}$$

Alt problemler

$$\mu^t = \arg \min_{\mu} \sum_{k=1}^K \sum_{i=1}^n \|x_i - \mu_k\|^2 \underbrace{z_{ik}^{t-1}}_{\text{sabit}} \rightarrow$$

$$\mu_k^t = \frac{\sum_{i=1}^n x_i z_{ik}^{t-1}}{\sum_{i=1}^n z_{ik}^{t-1}}$$

$\mu_k^t$  :  $t-1$  adımındaki  $k$  grubunun geometrik merkezi (centroid)

$$z^t = \arg \min_z \sum_{k=1}^K \sum_{i=1}^n \|x_i - \underbrace{\mu_k^t}_{\text{sabit}}\|^2 z_{ik}$$

$$\begin{aligned}\text{öyle ki } \sum_{k=1}^K z_{ik} &= 1, \quad i = 1, \dots, n; \\ z_{ik} &\in \{0, 1\}, \quad i = 1, \dots, n; k = 1, \dots, K.\end{aligned}$$



Eğer  $k = \arg \min \{\|x_i - \mu_k^t\| : k = 1, \dots, K\}$  ise  $z_{ik}^t = 1$

$i$  noktasını, en yakınındaki geometrik merkezin grubuna ata

# K-Ortalamalar Gruplandırma

## Yerel Çözümler

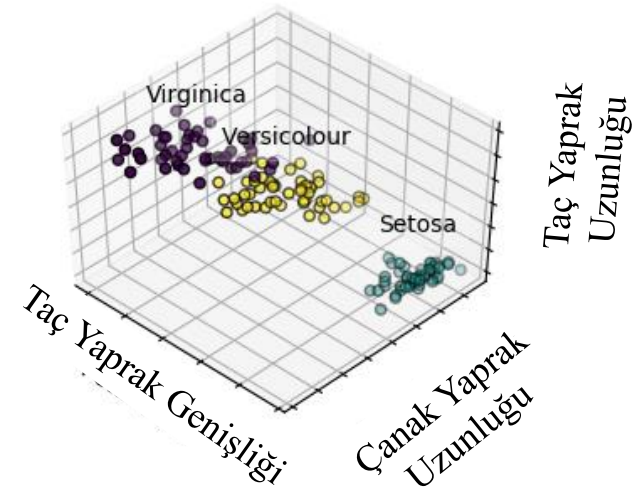




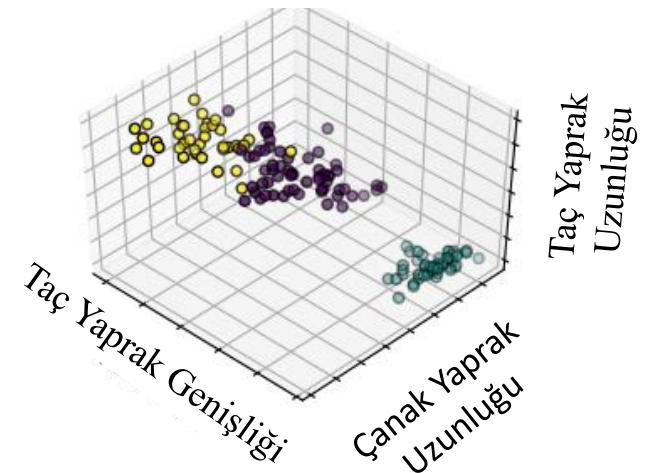
# Pratikte

```
1 # Modified from
2 # https://scikit-learn.org/stable/auto_examples/clu
3 # plot_cluster_iris.html#sphx-glr-auto-examples-clu
4
5 import numpy as np
6 import matplotlib.pyplot as plt
7 from mpl_toolkits.mplot3d import Axes3D
8
9 from sklearn.cluster import KMeans
10 from sklearn import datasets
11
12 iris = datasets.load_iris()
13 X = iris.data
14 y = iris.target
15
16 num_of_clusters = 3
17 k_means_iris = KMeans(n_clusters=num_of_clusters)
18
19 k_means_iris.fit(X)
20 labels = k_means_iris.labels_
21
22 # Plotting the cluster results
```

Gerçek Etiketler



K-Ortalama Etiketleri, K=3



# *K*-Medoit Gruplandırma Algoritması

- *K*-ortalama gibi çalışır fakat herhangi benzemezlik ölçüsüyle çalışır
- Grup merkezi gözlem noktalarından birisidir :

$$i_k^t = \arg \min_{\{i: C(i)=k\}} \sum_{C(j)=k} d(x_i, x_j)$$
$$\mu_k^t = x_{i_k^t}$$

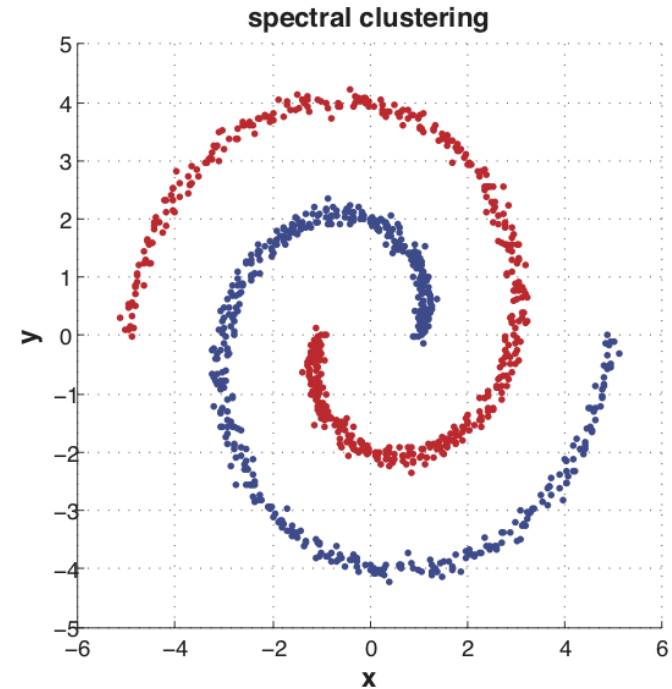
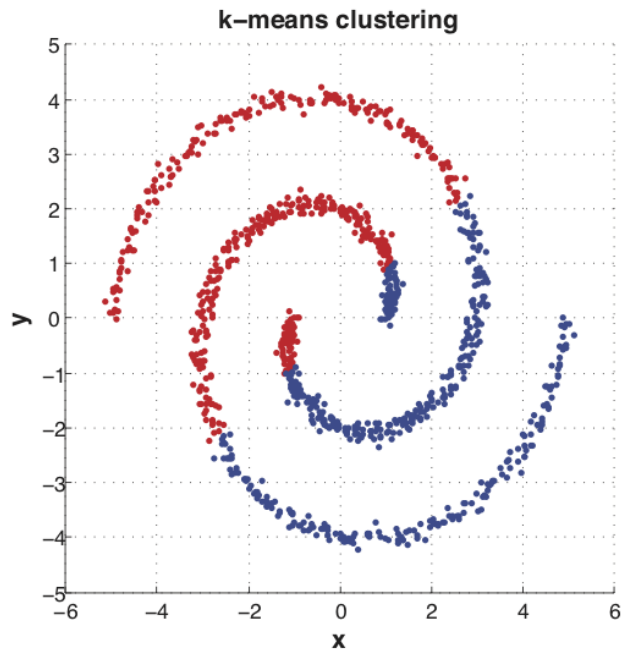
- Aynı şekilde, her bir gözlem noktası en yakındaki grup merkezine atanır:

$$C(i) = \arg \min_{k=1, \dots, K} d(x_i, \mu_k^t)$$

- Ortalamaları kullanmadığımızdan *K*-ortalama algoritmasından daha uzun sürer

# Spektral Gruplandırma

**Fikir:** Bir benzerlik matrisi ile veri noktalarından oluşan bir çizge (graph) tanımlamak ve çizge özdeğer (eigenvalue) analizi ile bağlantılı (connected) altçizgeleri (subgraph) belirlemek

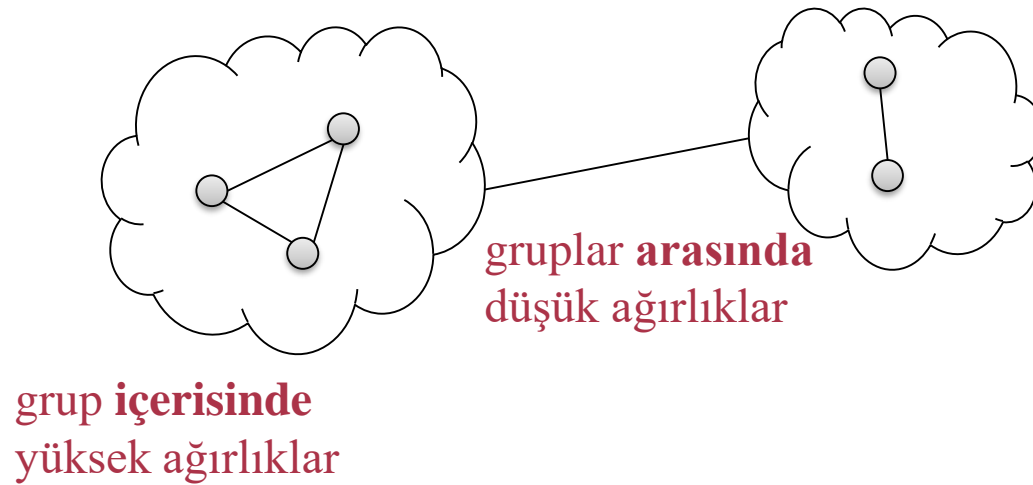


\*

# Spektral Gruplandırma

Benzerlik çizgesi – her veri noktası bir düğüm  $G = (V, E)$

Tamamen veya kısmen bağlı bir çizge olabilir (karşılıklı  $K$ -en yakın komşu)



$$w_{ij} = w_{ji} > \tau \geq 0$$



eşik değeri

$$w_{ij} = \exp \left( \frac{-\|x_i - x_j\|^2}{c} \right)$$

# Spektral Gruplandırma

$$W = [w_{ij}]_{i,j} \quad g_i = \sum_{j=1}^n w_{ij} \quad G = \text{diag}(g_1, g_2, \dots, g_n)$$

**Çizge Laplace  
Matrisi**

$$L = G - W$$

$$u = (u_1, u_2, \dots, u_n)^\top$$

$$u^\top L u = u^\top G u - u^\top W u = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} (u_i - u_j)^2$$

$$\begin{bmatrix} \blacksquare & & & \\ & \blacksquare & & \\ & & \ddots & \\ & & & \blacksquare \end{bmatrix}$$

*m* en küçük (=0) özdeğere karşılık gelen *m* özvektör bulunarak çizgenin *m* bileşeni (altçizge) belirlenir

# Spektral Gruplandırma

$$u \in \mathbb{R}^n$$

$$u^\top L u = u^\top G u - u^\top W u$$

$$= \sum_{i=1}^n g_i u_i^2 - \sum_{i,j=1}^n u_i u_j w_{ij}$$

$$= \frac{1}{2} \sum_{i,j=1}^n w_{ij} (u_i - u_j)^2$$

$w_{ij} \geq 0$

$$u^\top L u \geq 0$$

bütün özdeğerler  $\geq 0$

$$\underbrace{\begin{bmatrix} \sum_{j=1} w_{1j} & -w_{12} & \dots & -w_{1n} \\ -w_{21} & \sum_{j=1} w_{2j} & \dots & -w_{2n} \\ \vdots & \vdots & \dots & \vdots \\ -w_{n1} & -w_{n2} & \dots & \sum_{j=1} w_{nj} \end{bmatrix}}_L \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

en küçük  
özdeğer

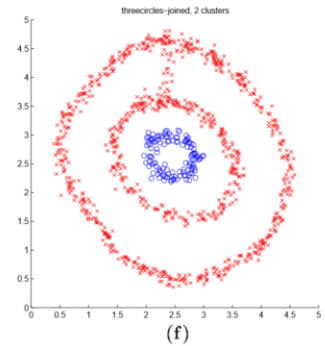
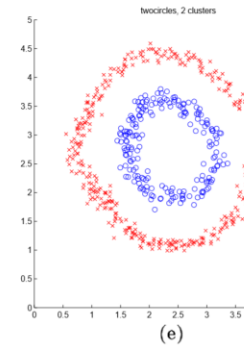
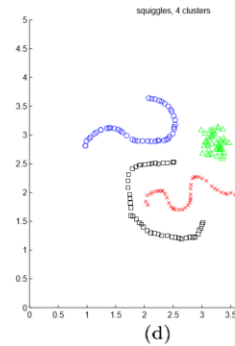
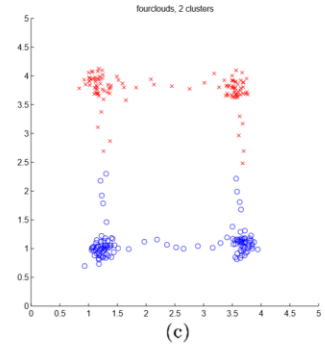
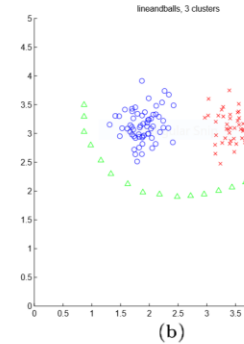
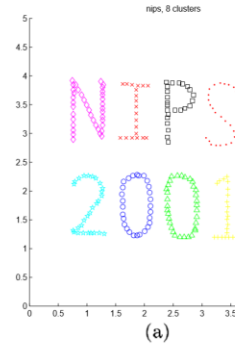
# Spektral Gruplandırma

$$L = G - W$$

$m$  en küçük  
özdeğere karşılık  
gelen  $m$  özvektör

$$Z = \begin{bmatrix} | & | & \dots & | \\ z_1 & z_2 & \dots & z_m \\ | & | & \dots & | \end{bmatrix}_{n \times m}$$

K-Ortalamlar  
Gruplandırma



# Hiyerarşik Gruplandırma

**Fikir:** Mevcut iki grup arasındaki benzerliğe (ya da benzemezliğe) bakarak yeni gruplar oluşturma

Sırasıyla  $m_1$  ve  $m_2$  noktadan oluşan  $C_1$  ve  $C_2$  grupları

Tek Bağlamalı (*yığinsal* kümeleme)

$$d_{\text{SL}}(C_1|C_2) = \min_{x_i \in C_1, x_j \in C_2} d(x_i, x_j)$$

Tam Bağlamalı

$$d_{\text{CL}}(C_1|C_2) = \max_{x_i \in C_1, x_j \in C_2} d(x_i, x_j)$$

Grup Ortalaması

$$d_{\text{GA}}(C_1|C_2) = \frac{1}{m_1 m_2} \sum_{x_i \in C_1} \sum_{x_j \in C_2} d(x_i, x_j)$$



# Hiyerarşik Gruplandırma

**Gruplar:**  $\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}$

$$s(\{x_1\}|\{x_2\}) = 5.0$$

$$s(\{x_1\}|\{x_3\}) = 0.5$$

$$s(\{x_1\}|\{x_4\}) = 4.5$$

$$s(\{x_1\}|\{x_5\}) = 2.0$$

$$s(\{x_2\}|\{x_3\}) = 4.7$$

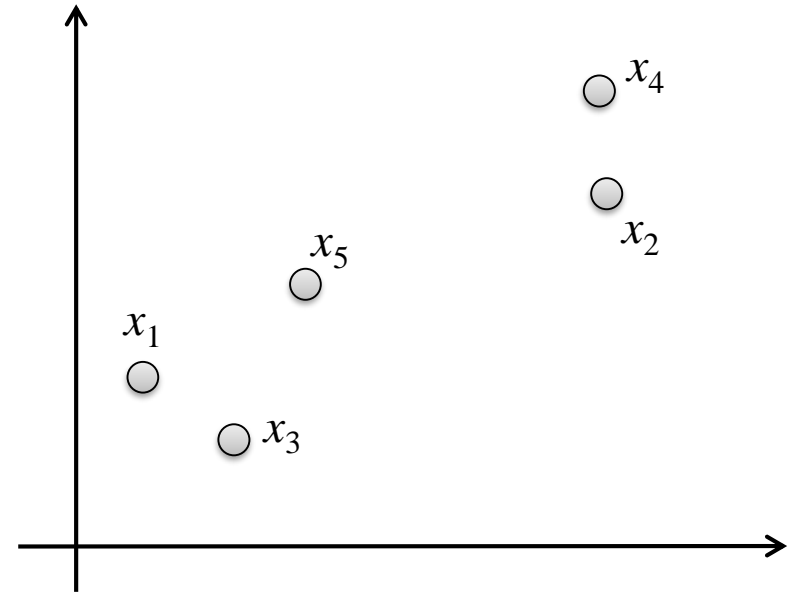
$$s(\{x_2\}|\{x_4\}) = 0.6$$

$$s(\{x_2\}|\{x_5\}) = 3.0$$

$$s(\{x_3\}|\{x_4\}) = 4.0$$

$$s(\{x_3\}|\{x_5\}) = 2.2$$

$$s(\{x_4\}|\{x_5\}) = 2.5$$



**Gruplar:**  $\{x_1, x_3\}, \{x_2\}, \{x_4\}, \{x_5\}$

$$s(\{x_1, x_3\}|\{x_2\}) = 5.0$$

$$s(\{x_1, x_3\}|\{x_4\}) = 4.5$$

$$s(\{x_1, x_3\}|\{x_5\}) = 2.2$$

$$s(\{x_2\}|\{x_4\}) = 0.6$$

$$s(\{x_2\}|\{x_5\}) = 3.0$$

$$s(\{x_4\}|\{x_5\}) = 2.5$$

**Gruplar:**  $\{x_1, x_3\}, \{x_2, x_4\}, \{x_5\}$

$$s(\{x_1, x_3\}|\{x_2, x_4\}) = 5.0$$

$$s(\{x_1, x_3\}|\{x_5\}) = 2.2$$

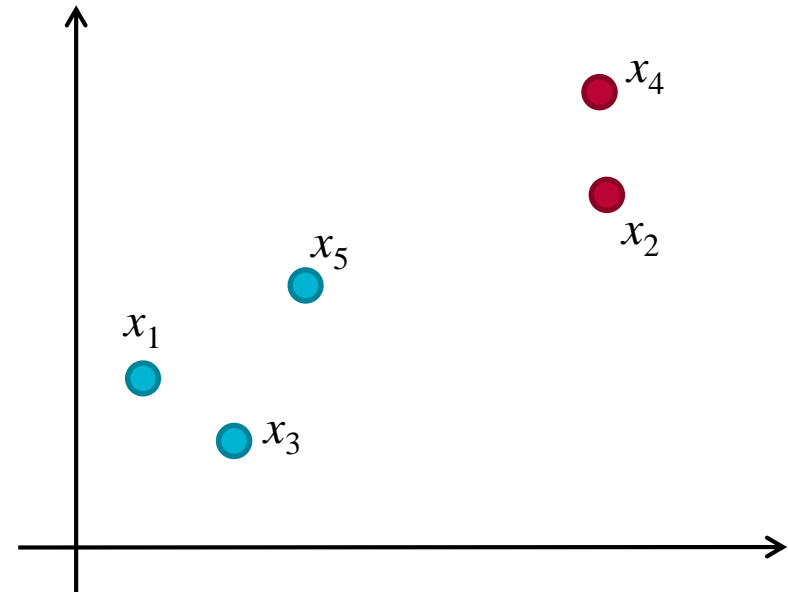
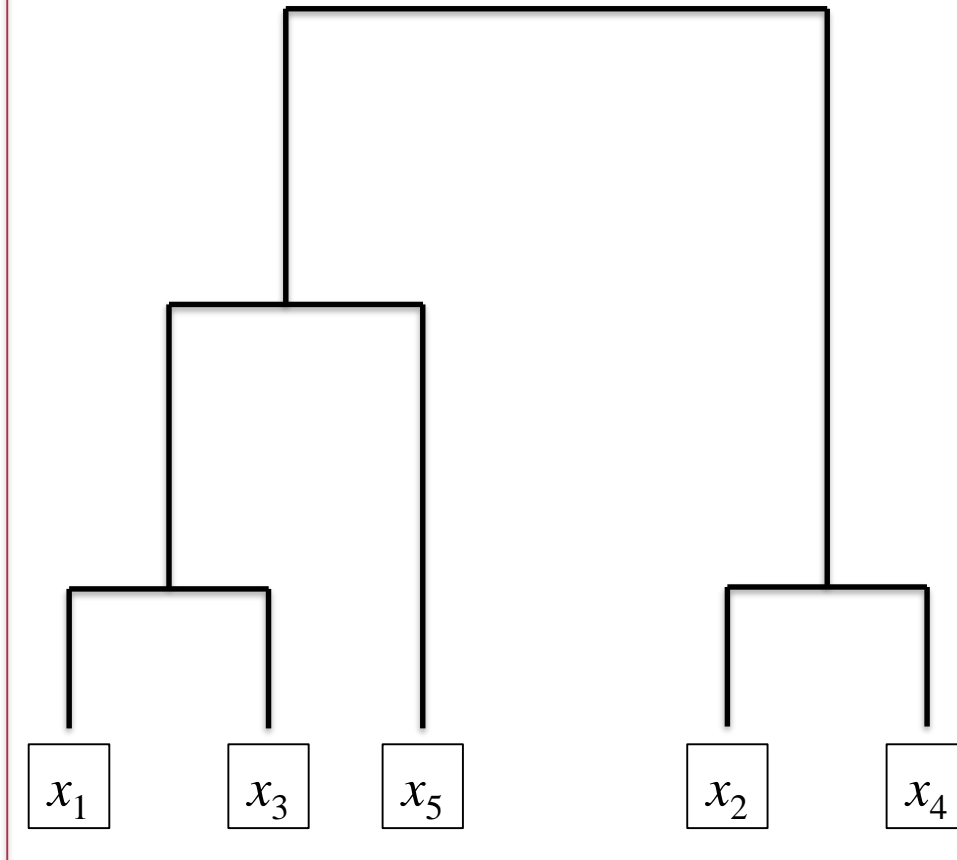
$$s(\{x_2, x_4\}|\{x_5\}) = 3.0$$

**Gruplar:**  $\{x_1, x_3, x_5\}, \{x_2, x_4\}$

$$s(\{x_1, x_3, x_5\}|\{x_2, x_4\}) = 5.0$$

# Hiyerarşik Gruplandırma

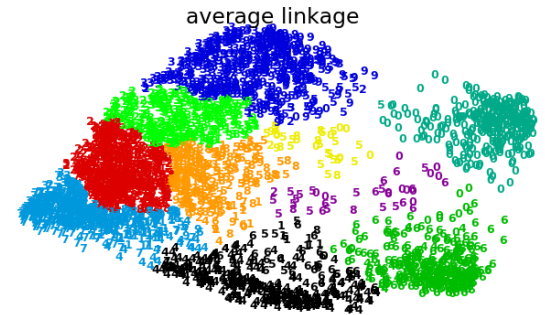
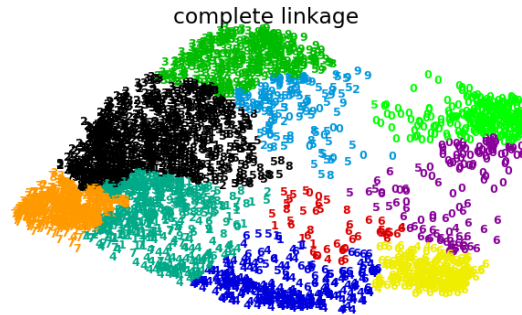
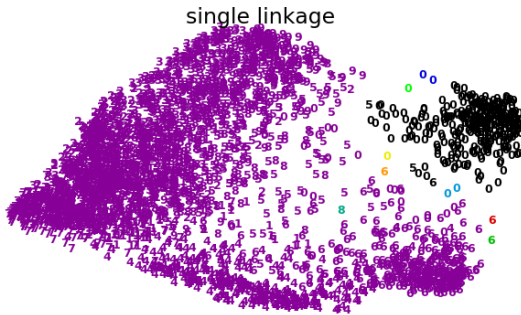
Ağaç Diyagramı  
(Dendrogram)



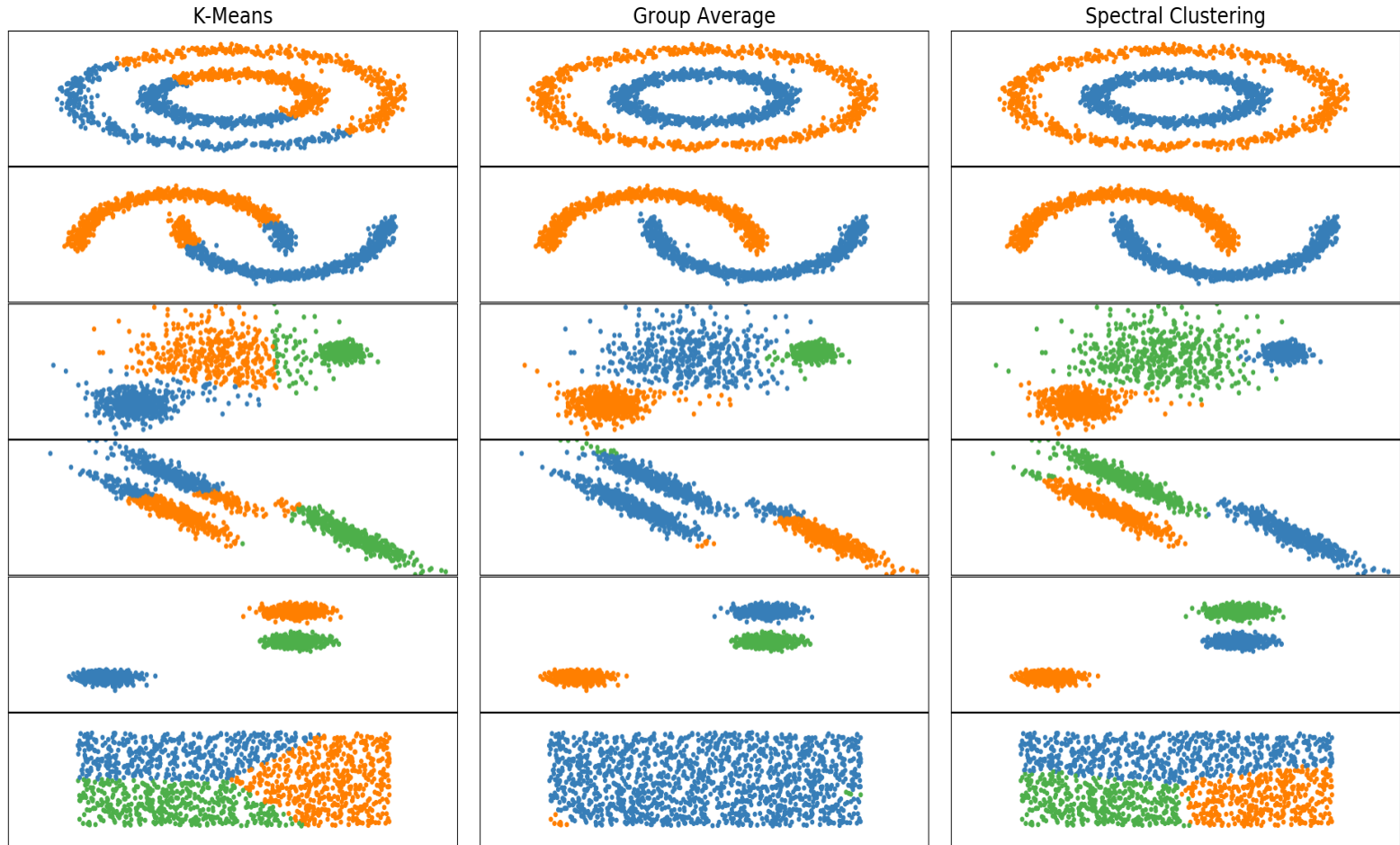
Eğer **üç** grubumuz olmasına karar verirsek, bunlar:  
 $\{x_1, x_3\}, \{x_5\}$  ve  $\{x_2, x_4\}$

# Hiyerarşik Gruplandırma

- Tek bağlamalı gruplandırma, diğer gözlemler epey farklı olmasına rağmen iki grubu daha az gözlem uğruna birleştirebilir
- Tam bağlamalı gruplandırma bir gözlemi kendisine çok benzeyen gözlemleri içeren gruplara atamaktansa yanlış bir gruba atayabilir
- Grup ortalaması kullanarak gruplandırma ikisi arasındadır, fakat sayısal ölçeğe bağımlıdır (ölçek değişmezliği söz konusu değildir)



# Karşılaştırma



(kaynak)

# Matris Ayırıştırma

- Veri ayrık bir matris olarak verilmiş ( $Y$ )
- Amaç  $Y$  matrisini iki matris olarak ayırıştırma
- Klasik örnek film tavsiye sistemleri

	M1	M2	M3	M4
Amy	5	2	?	2
Brad	4	?	?	3
Carl	1	1	?	4
Dan	2	?	4	5
Emily	?	2	?	4

# Matris Ayırıştırma

	M1	M2	M3	M4
Amy	5	2	?	2
Brad	4	?	?	3
Carl	1	1	?	4
Dan	2	?	4	5
Emily	?	2	?	4

$$\underbrace{\begin{bmatrix} 5 & 2 & ? & 2 \\ 4 & ? & ? & 3 \\ 1 & 1 & ? & 4 \\ 2 & ? & 4 & 5 \\ ? & 2 & ? & 4 \end{bmatrix}}_Y \approx \underbrace{\begin{bmatrix} * & * \\ * & * \\ * & * \\ * & * \\ * & * \end{bmatrix}}_{X_U} \underbrace{\begin{bmatrix} * & * & * & * \\ * & * & * & * \end{bmatrix}}_{X_M}$$

Gizli boyut (sabit)

# Matris Ayırıştırma

$$(X_U^*, X_M^*) = \arg \min_{X_U, X_M} \frac{1}{n} \|Y - X_U X_M\|_F^2$$

- Toplam reyting sayısı,  $n$
- Birçok düzenleyici terim eklenebilir
- Gizli boyut problemin çözümünden önce sabit alınır
- Ayırıştırılanlar grup gibi kullanılabilir
- Dışbükey olmayan bir problem

# Pratike

```
1 import numpy as np
2 from sklearn.decomposition import NMF
3
4 Y = np.array([[5, 2, 0, 2],
5               [4, 0, 0, 3],
6               [1, 1, 0, 4],
7               [2, 0, 4, 5],
8               [0, 2, 0, 4]])
9
10 model = NMF(n_components=2, init='random', random_state=0)
11
12 X_U = model.fit_transform(Y)
13 X_M = model.components_
14
15 Y_pred = np.dot(X_U, X_M)
16
17 np.set_printoptions(precision=2)
18 print(np.hstack((Y, Y_pred)))
```

$$\underbrace{\begin{bmatrix} 5 & 2 & ? & 2 \\ 4 & ? & ? & 3 \\ 1 & 1 & ? & 4 \\ 2 & ? & 4 & 5 \\ ? & 2 & ? & 4 \end{bmatrix}}_Y \approx \underbrace{\begin{bmatrix} * & * \\ * & * \\ * & * \\ * & * \\ * & * \end{bmatrix}}_{X_U} \underbrace{\begin{bmatrix} * & * & * & * \\ * & * & * & * \end{bmatrix}}_{X_M}$$

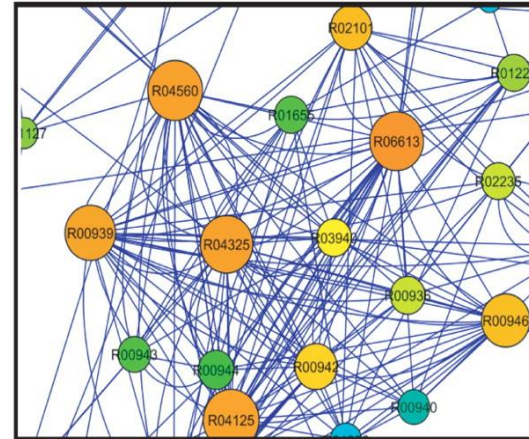
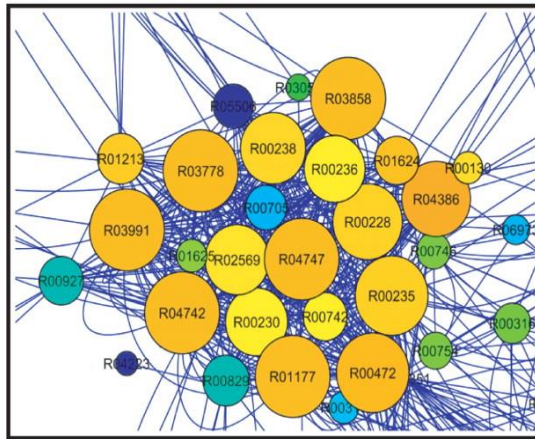
```
[9] print(np.hstack((Y, Y_pred)))...
```

```
[[5.  2.  0.  2. | 5.09 1.18 0.  2.23]
 [4.  0.  0.  3. | 3.9  1.02 0.43 2.68]
 [1.  1.  0.  4. | 1.31 0.66 1.28 3.48]
 [2.  0.  4.  5. | 1.48 0.96 2.22 5.67]
 [0.  2.  0.  4. | 0.54 0.53 1.46 3.54]]
```



# Sayfa Sıralama (PageRank) Algoritması

- Eğer bir sayfa diğer önemli sayfalarca atfediliyorsa, önemlidir
- Eğer atfedilen sayfa, başka sayfalara az atıfta bulunuyorsa önemi artar
- Google bu algoritma ile başladı (şu an daha karmaşık bir algoritma var)
- Başka alanlarda da kullanılıyor



\* *When the Web meets the cell: using personalized PageRank for analyzing protein interaction networks*, G. Iván, V. Grolmusz, Bioinformatics, (27) 3, 405-407, 2011.

# Sayfa Sıralama (PageRank) Algoritması

- Sıralanacak  $n$  sayfa
  - $L_{ij} = 1$ , eğer sayfa  $j$  sayfa  $i$ 'ye atıfta bulunuyorsa
- $$c_j = \sum_{i=1}^n L_{ij}$$

sayfa  $i$ 'nin  
sıralaması

$$p_i = (1 - d) + d \sum_{j=1}^n \left( \frac{L_{ij}}{c_j} \right) p_j \quad (d \approx 0.85)$$

$$\mathbf{D}_c = \text{diag}(\mathbf{c})$$

$$\mathbf{p} = (1 - d)\mathbf{e} + d\mathbf{L}\mathbf{D}_c^{-1}\mathbf{p}$$

$$\mathbf{e}^\top \mathbf{p} = n$$

$$\mathbf{p} = \underbrace{\left( \frac{(1 - d)}{n} \mathbf{e}\mathbf{e}^\top + d\mathbf{L}\mathbf{D}_c^{-1} \right)}_{\mathbf{A}} \mathbf{p}$$

- Güç Yöntemi:  
(Power method)

$$\begin{aligned} \mathbf{p}_k &\leftarrow \mathbf{A}\mathbf{p}_{k-1} \\ \mathbf{p}_k &\leftarrow n \frac{\mathbf{p}_k}{\mathbf{e}^\top \mathbf{p}_k} \end{aligned}$$

# Özet

- Ana Bileşenler Analizi
- K-Ortalamalar Gruplandırma
- Hiyerarşik Gruplandırma
- Spektral Gruplandırma
- Matrix Ayrıştırma
- Sayfa Sıralama (PageRank) Algoritması