Clustering

GMM : already studied. : Statistical $P(X)$

Optimization POV.

K-means and K-medoids

A : Data $N \times p$

Matrix $_{N \times N}$

B : Similarity metric $\| \ \|_2$

$\mathbb{R}^p$

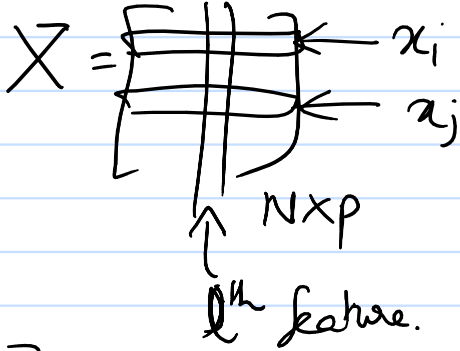
dissimilarity (x_i, x_j)

eg 1

$$= \sum_{k=1}^{p} (x_{ik} - x_{jk})^2$$

eg 2 :

$$\rho = \frac{\sum_{k=1}^{p} x_{ik} x_{jk}}{\|x_i\|_2 \, \|x_j\|_2}$$



$$\rightarrow \quad \rho = \frac{E\left[(X - E[X])(Z - E(Z))\right]}{\sqrt{\mathrm{Var}(X) \cdot \underbrace{\mathrm{Var}(Z)}_{E\left[(Z - E(Z))^2\right]}}}$$

$$\underbrace{\cancel{\|x_i - x_j\|_2^2}} = 2(1 - \rho) \quad \text{if} \quad \|x_i\|_2 = 1$$

$$\|x_j\|_2 = 1$$

$\mathbb{R}^p$

K means:

$1 \longrightarrow \{1, \ldots, K\}$

$\vdots$

$i \longrightarrow \{1, \ldots, K\}$

$\vdots$

$N \longrightarrow \{1, \ldots, K\}$

$$\boxed{C : [N] \to [K]}$$

$$f\,\text{Scoring}(\underline{\underline{C}}) = \frac{1}{2} \sum_{k=1}^{K} \underline{\underline{\sum_{i: C(i)=k} \sum_{j: C(j)=k}}} \underline{\underline{d(x_i, x_j)}}$$

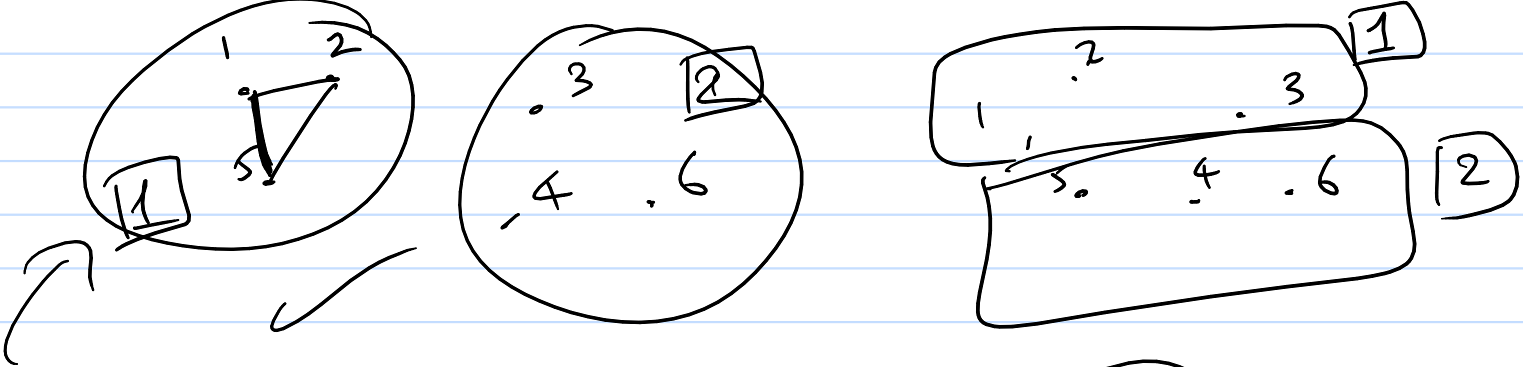
$\{1, 2, 5\} \rightarrow \underline{1}^{st}$ cluster

$d(1,2) + d(2,5) + d(1,5)$

$d(2,1) + d(5,2) + d(5,1)$

$$\boxed{\min_{C} \text{Scoring}(C)}$$

<u>Computationally hard.</u>



$$\text{Scoring}'(C) = \sum_{k=1}^{K} N_k(C) \cdot \sum_{i:C(i)=k} \| x_i - \bar{x}_k \|_2^2$$

$$(C)$$

$$\bar{x}_k = \frac{1}{N_k} \sum_{i:C(i)=k} x_i$$

$$\text{Relaxed Scoring}(C, \{m_k\}) = \sum_{k=1}^{K} N_k \cdot \sum_{i: C(i)=k} \|x_i - m_k\|_2^2$$

$x_i \in \mathbb{R}^p$, $m_k \in \mathbb{R}^p$

for iter = 1, ..., 1000

given $\{m_k\}$, Solve for C ☆

given C, Solve for $\{m_k\}$ →

"K-means"

$$C^{(0)} \to \{m_k\}^{(0)} \to C^{(1)} \to \{m_k\}^{(1)}$$

K-medoids

Data

Sim/dissimilarity measure

$= d(x_i, x_j)$

$$D = \begin{bmatrix} & \vdots & \\ \cdots & d(x_i, x_j) & \\ & & \end{bmatrix}_{N \times N}$$

x x x cat x x x x

x x x dog x x x x

x x x fox x x x x

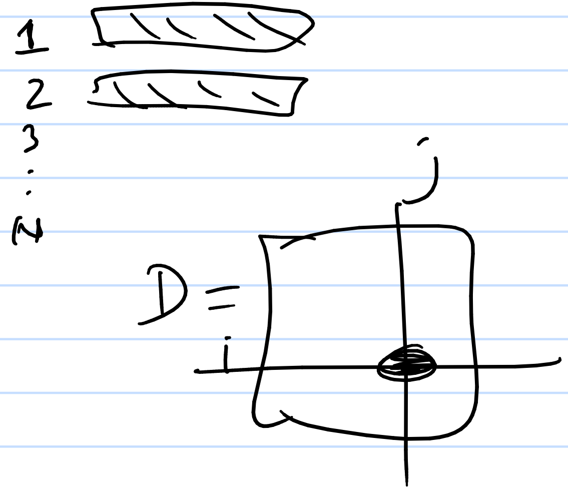
$l = [0 \; 1 \; 10]$

$\{m_k\} \longrightarrow \{i_k\}$

for

$C \longrightarrow \{i_k\}$

$\{i_k\} \longrightarrow C^{new}$

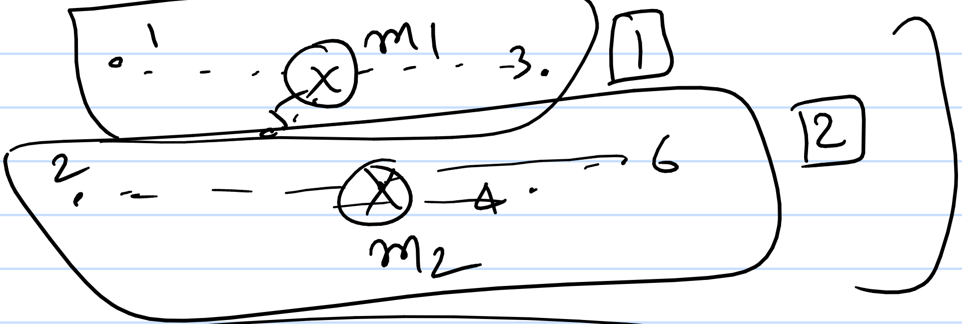


1st cluster

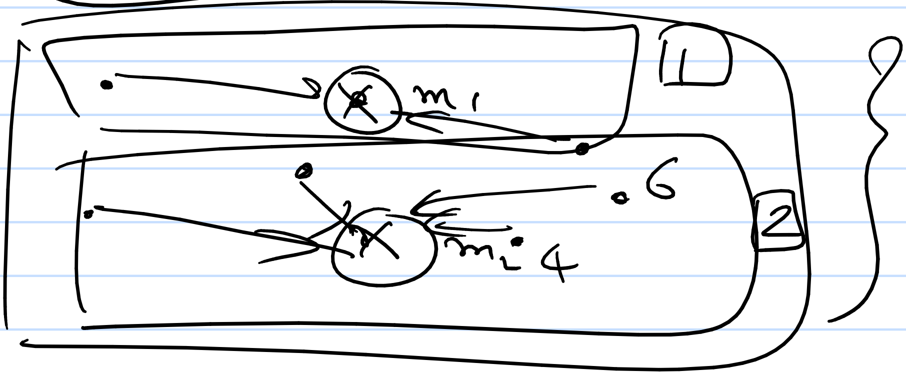
1 2 5

$$\min_{i:\,\text{remaining pts}} \sum d(\text{index}, i)$$

index = 1 or 2 or 5

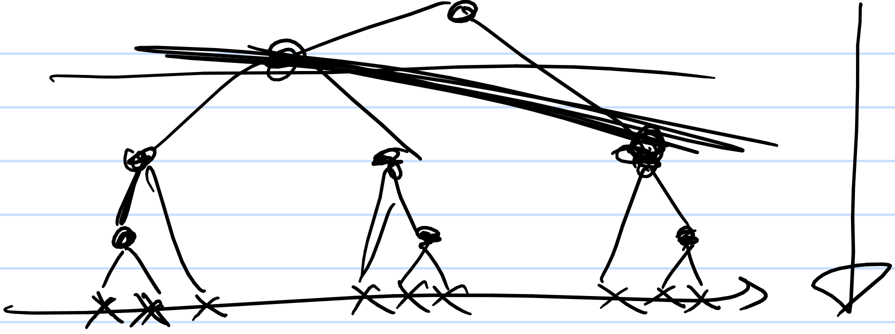


$C \to \{m_k\}$



$\{m_k\} \to C$

Hierarchical clustering



agglomerative

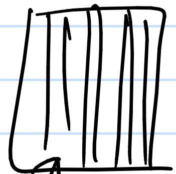
divisive

Mixed membership clustering

Principal Components

linear

$$X_{N\times p} \longrightarrow X'_{N\times q}$$



$$x_i \longrightarrow \lambda_i$$

$x_i \in \mathbb{R}^p$, $\lambda_i \in \mathbb{R}^q$

q orthogonal rows $\rightarrow (V_q^T)_{q\times p}$

$$x_i \xrightarrow{\;V_q^T\;}_{q\times p} \lambda_i \xrightarrow{\;V_q\;}_{p\times q} \tilde{x}_i$$

$$q \qquad\qquad p$$

$$\text{obj} \quad \boxed{\min_{V_q} \sum_{i=1}^{N} \| x_i - \underbrace{V_q V_q^T}_{p\times p} \underbrace{x_i}_{\lambda_i} \|_2^2}$$

x

Non convex opt problem.

$$U, D, V^T \leftarrow \text{svd}(X)$$

$$(V_q)_{p \times q}$$

$$\begin{bmatrix} \vdots & & \vdots \\ v_i & \cdots & v_j \\ \vdots & & \vdots \end{bmatrix}$$

$$v_i^T v_j = 0$$

$$\underbrace{U}_{N \times p},\ \underbrace{D}_{p \times p},\ \underbrace{V^T}_{p \times p} \leftarrow \underline{\underline{svd}}(X)_{N \times p}$$

V_q

$p \times p$

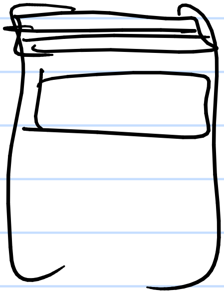
$$D = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_p \end{bmatrix}$$

$$|\sigma_1| > |\sigma_2| > \dots |\sigma_p|$$

$$\underbrace{V_q}_{p\times q} = \begin{bmatrix} | & & | \\ V_{(1)} & \cdots & V_{(q)} \\ | & & | \end{bmatrix}$$

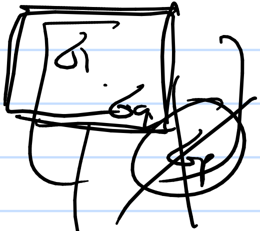
principal component directions

$$\underbrace{U D_q}_{N\times q} =$$



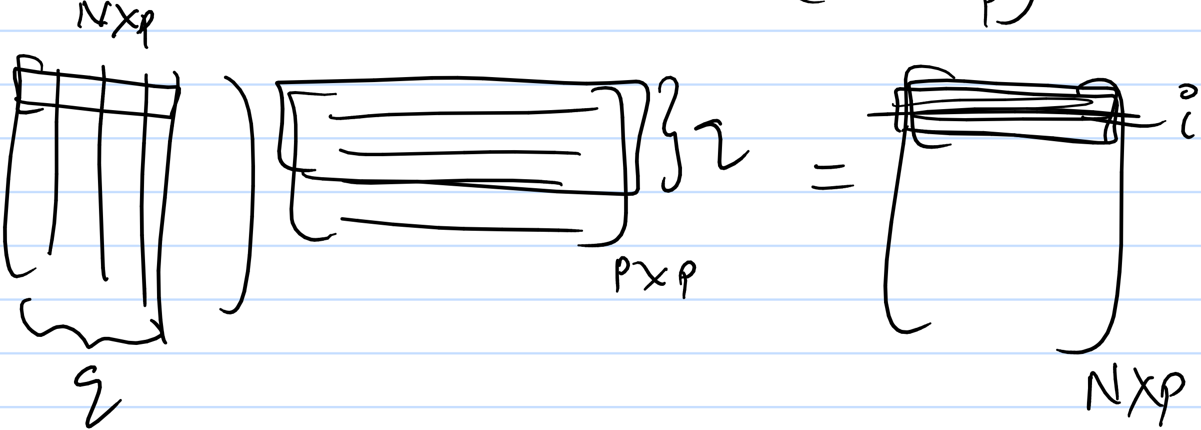
$$\underset{N\times p}{U} \cdot \underset{p\times p}{D} \cdot \underset{p\times p}{V^T} = X_{N\times p}$$

$$\Downarrow$$



$$U D_q V_q^T$$

$$\underbrace{UD}V^T = X_{N\times p} \qquad \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_p \end{bmatrix}$$



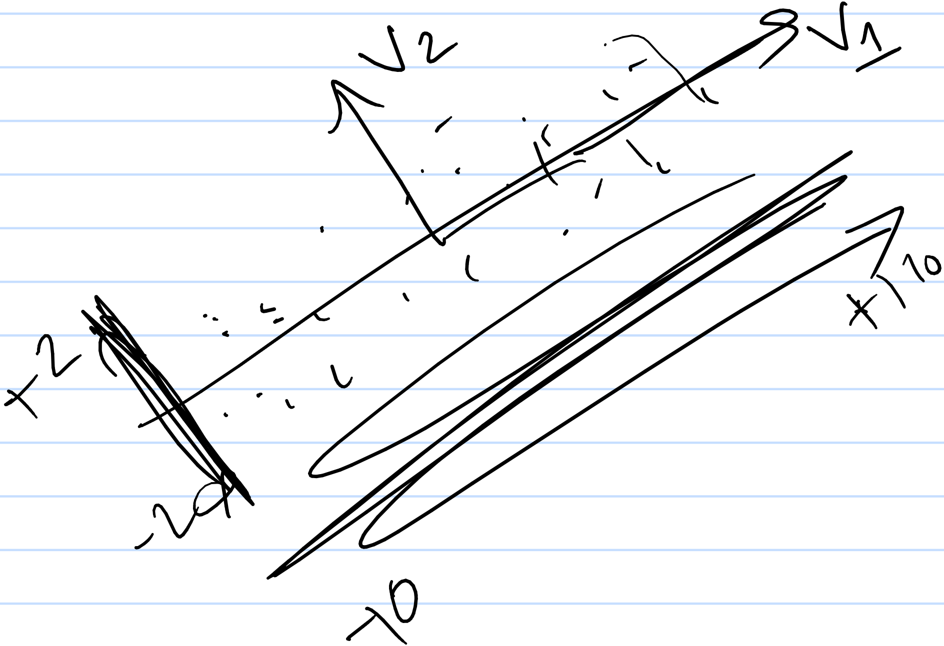
$$[u_{i1}d_1, \dots, u_{iq}d_q]$$

$$\underbrace{\begin{bmatrix} V_{(1)} & \cdots & V_{(q)} \end{bmatrix}}_{p \times q} \underbrace{\underbrace{\begin{bmatrix} u_{i1} d_1 \\ \vdots \\ u_{iq} d_q \end{bmatrix}}_{q \times 1}}_{\lambda_i} \approx \underbrace{\begin{bmatrix} \\ \\ \\ \end{bmatrix}}_{p \times 1} \; x_i$$

$$Ax = b$$

$$\begin{bmatrix} A_1 & A_2 & A_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = b$$

$$x_1 A_1 + x_2 A_2 + x_3 A_3 = b$$


v2
v1

$$X_{N\times p} V_{(1)\,p\times 1} = Z_{N\times 1}$$

$D^T = D$

$Var(z)$ is high

$$\underline{Var(z)} = \frac{1}{N} z^T z = \frac{1}{N} V_{(1)}^T \underbrace{X^T X}_{} V_{(1)}$$

$$= \boxed{V D^2 V^T}$$

$$\underbrace{V D U^T}_{X^T} \; \underbrace{U D V^T}_{X}$$

$$X^T X = \sum_{i=1}^{p} \underline{\underline{d_i^2}} \, v_{(i)} v_{(i)}^T \quad p \times p$$

$$X^T X = \underline{\underline{V D^2 V^T}}$$

$$D = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_p \end{bmatrix}$$

$$\mathrm{Var}(z) = \frac{1}{N} v_{(1)}^T \left(\quad \right) v_{(1)}$$

$$= \frac{1}{N} \sum_{i=1}^{p} d_i^2 \left(\underbrace{v_{(1)}^T v_{(i)}} \; \underbrace{v_{(i)}^T v_{(1)}} \right)$$

$$= \frac{d_1^2}{N}$$

Spectral clustering

$$D_{N\times N} \longrightarrow W_{N\times N} \longrightarrow \boxed{?} \longrightarrow X^{new}_{N\times q}$$

$\longrightarrow$ K-means on

on "this" data

$$W_{N\times N} \longrightarrow L_{N\times N} = G - W$$

diagonal $\sum_{i=1}^{N} w_{\cdot i}$

$N \times N$

(1) Operational

eigenvalues, eigenvectors (L)

$0 = \lambda_1, V_1 = \mathbf{1}_{N\times 1}$

$0.1 = \lambda_2, V_2$

$0.2 = \lambda_3, V_3$

$= X^{new}$

$N \times 2$

(2) Interpretation

$$L = \begin{array}{c} r \\ bl \\ gr \end{array} \begin{bmatrix} \;\; & 0 & 0 \\ 0 & \;\; & 0 \\ 0 & 0 & \;\; \end{bmatrix}$$

eigenvalues, eigenvectors

$$\mathbf{0} \qquad \begin{array}{c} r \\ \\ \\ bl \\ \\ gr \\ \\ \end{array} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \; & \; & \; \end{bmatrix}_{\cancel{N} \times p}$$