Clustering

① dissimilarity

$\{x_i\}_{i=1}^{N}$ $\quad x_i \in \mathbb{R}^p$

$$d_{ii'} = \underbrace{D(x_i, x_{i'})}_{= \sum_{j=1}^{p} d_j(x_{ij}, x_{i'j})}$$

$$d_j(x_{ij}, x_{i'j}) \rightarrow (x_{ij} - x_{i'j})^2$$

$$D_{N\times N} = \begin{bmatrix} \quad \end{bmatrix}_{N\times N}$$

$\underline{\underline{p=3}}$

$$\underset{a}{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}} \quad \underset{b}{\begin{bmatrix} -2 \\ 1 \\ 5 \end{bmatrix}} = \frac{a^T b}{\|a\|_2 \|b\|_2}$$

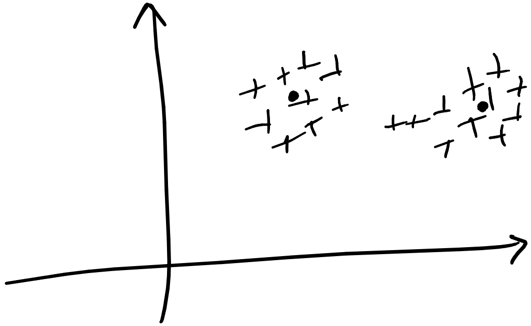
$d_{ii'}$

$$s_{ii'} = \frac{1}{d_{ii'}}$$

$$= 1 - d_{ii'}$$

$$= e^{-d_{ii'}}$$

1. Mixture models : GMM. P_X



2. Optimization formulations.

$C(i) \qquad i \in 1, \dots, N$

$C(i) \in \{1, 2, 3\}$

$$\min_{C} \; W(C) = \frac{1}{2} \sum_{k=1}^{K} \sum_{i:C(i)=k} \sum_{i':C(i')=k} d(x_i, x_{i'}) \Big]$$

each Coordinate is <u>Continuous</u>.

$$d_{ii'} = \|x_i - x_{i'}\|_2^2$$

$$W(C) = \sum_{K=1}^{K} N_K \cdot \sum_{i:C(i)=K} \|x_i - \bar{x}_K\|_2^2$$

$\bar{x}_K$ = mean of K^{th} cluster.

$\in \mathbb{R}^p$

$$\bar{x}_K = \underset{m \in \mathbb{R}^p}{\operatorname{argmin}} \|x_i - m\|_2^2$$

$$W\left(C, \{m_k\}_{k=1}^{K}\right) = \sum_{k=1}^{N} N_k \sum_{i:C(i)=k} \| x_i - m_k \|_2^2$$

Key idea: alternatively minimize over C and $\{\underline{\underline{m_k}}\}_{k=1}^{K}$

<u>EM Relation</u>:

$$\underline{\underline{C(i)}} \longleftrightarrow \underline{\underline{q(\underline{\underline{z_i}}=1 \mid x_i)}}\ \star$$

$$\{m_k\} \qquad\qquad \{\mu_k\ \Sigma_k,\ \pi_k\}$$

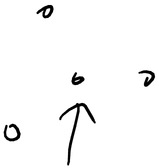
K-medoids

$D_{N \times N}$

for $j = 1 \ldots T$

① given $\{C(i)\}_{i=1}^{N}$: $\quad i_k = \underset{i:\, C(i)=k}{\arg\min} \sum_{\substack{i':\\ C(i')=k}} d_{ii'}$

② Given $\{i_k\}$:



for $i = 1, \ldots, N$

$$C(i) = \underset{k=1 \ldots K}{\arg\min}\; d_{i i_k}$$

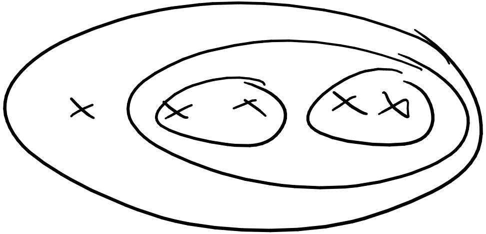
Hierarchical Clustering : dendrogram : tree

$$d(G, H) \leftarrow \min_{i \in G,\ i' \in H} d_{ii'}$$

$\{1, 2, 3, 4\}$ $\{7, 10, 12\}$

- agglomerative

- divisive



Principal Components $\quad \mathbb{R}^p \ni x_i \quad i=1 \ldots N$

$\Downarrow$

$q < p. \quad \mathbb{R}^q \ni \lambda_i$

$$A x_i = \lambda_i$$

A: $q \times p$

$$\tilde{x}_i = \underbrace{V_q}_{p \times q} \underbrace{(V_q^T)}_{q \times p \text{ matrix}} x_i$$

find V_q : $\tilde{x}_i \approx x_i$

$$\min_{V_q} \sum_{i=1}^{N} \| x_i - V_q V_q^T x_i \|_2^2$$

SVD will get us V_q : $\underset{p \times q}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}$

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

$|d_1| > |d_2| \ldots$

$U_{N \times p}$

$$D_{p \times p} = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_p \end{bmatrix}$$

$$V = \begin{array}{c} v_1 \; v_2 \; \ldots \; v_q \\ \left[\begin{array}{c|c|c|c} & & & \\ & & & \\ & & & \end{array}\right]_{p \times p} \end{array}$$

$$\underset{N \times p}{UD} = \begin{array}{c} \quad\quad\quad q \\ \begin{bmatrix} \lambda_1 & \\ \lambda_2 & \\ & \\ & \end{bmatrix} \end{array}$$

$$\mathrm{Var}\big(\underbrace{X}_{N\times p}\underbrace{v_1}_{p\times 1}\big) = \frac{d_1^2}{N} \quad \text{is the largest.}$$

$$X^T X = \sum_{i=1}^{N} d_i^2 v_i \cdot v_i^T$$

Spectral Clustering. $\{x_i\}_{i=1}^{N} \longrightarrow \underline{\underline{d_{ii'}}} \longrightarrow s_{ii'}$

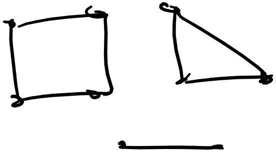
Graph. (undirected) V, E, W

V: N

W: weights on the edges.

① Laplacian matrix

$$L_{N\times N} = \begin{bmatrix} \sum_i w_{1i} & & 0 \\ & \ddots & \\ 0 & & \ddots \end{bmatrix} - W$$



② find eigenvectors $\underline{Z}_{N\times \underline{m}}$ dimensional

m smallest eigenvalues of L.

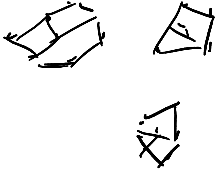
$$\text{svd}(L) = V D^2 V^T$$

$$\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix}$$

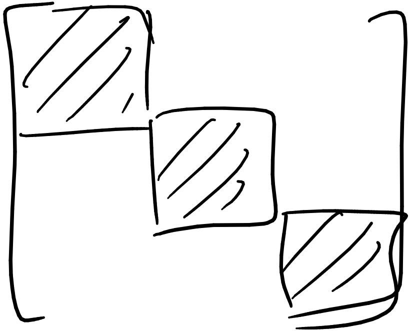
L : p-s-d matrix.

$$x^T L x \geq 0$$

③ Apply k-means on $Z_{N\times m}$.

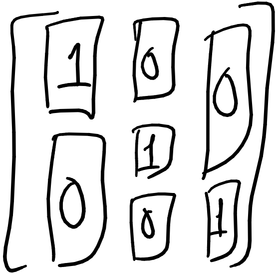


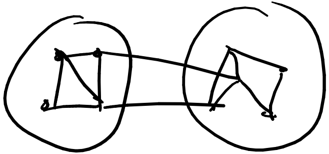
$W =$



$N \times N$

$L \longrightarrow \lambda_1 = 0 \quad \lambda_2 = 0 \quad \lambda_3 = 0$





Approximation to the normalized cut problem