

k-means

①

$$\min_{\{y_i\}, \{c_i\}} \sum_{j=1}^k \sum_{y_i=j} \|x_i - c_j\|^2$$

$$(x_i - c_j)^T (x_i - c_j) = x_i^T x_i - x_i^T c_j - c_j^T x_i + c_j^T c_j$$

$$\sum_j \sum_i \underbrace{x_i^T x_i - x_i^T c_j - c_j^T x_i + c_j^T c_j}_{\text{doesn't matter}}$$

$$\min \sum_j \sum_i -2 x_i^T c_j + \sum_j \sum_i c_j^T c_j$$

$$\psi_1, \psi_2, \dots, \psi_n \quad \psi_i \in \{1, \dots, n\}$$

$$\cup \psi_i = \{1, \dots, n\}$$

$$\psi_i \cap \psi_j = \emptyset$$

$$\rightarrow |\psi_j|$$

$$\min_{\psi, c} \sum_{j=1}^n l_j c_j^T c_j - 2 \sum_j \sum_{i \in \psi_j} x_i^T c_j$$

$$\rightarrow \frac{1}{|\psi_j|} \sum_{i \in \psi_j} x_i$$

$$\min_{\psi} - \sum_j l_j \sum_{i, s \in \psi_j} x_i^T x_s$$

$$\max_{\{\psi_i\}} \sum_{j=1}^n \frac{1}{l_j} \sum_{i, s \in \psi_j} x_i^T x_s$$

$$x_i \rightarrow \phi(x_i)$$

$$\phi(x_i)^T \phi(x_j)$$

$$\rightarrow k(x_i, x_j)$$

kernel
Trick

kernel matrix: $K = (k(x_i, x_j))_{n \times n}$ (2)
 symmetric, positive semi-definite

$$F = \left(F_{ij} = \frac{1}{l_i} \right)_{i \in (i,j) \in \Psi_K}$$

$$F_{ij} = 0 \text{ otherwise}$$

$$F = \begin{pmatrix} \# & & \\ & F_2 & \\ & & \ddots & \\ & & & F_n \end{pmatrix} \quad \text{Block Diagonal}$$

$$F_n = \frac{1}{l_n} \mathbf{1} \cdot \mathbf{1}^T$$

$$\boxed{\max_F \sum_{i,j=1}^n k_{ij} F_{ij} = \text{Tr}(KF)} \quad |$$

$$G_{ij} = \begin{cases} \frac{1}{\sqrt{l_j}} & \text{if } i \in \Psi_j \\ 0 & \text{otherwise} \end{cases}$$

$$G = \begin{pmatrix} g_1 & \dots & g_n \end{pmatrix} \quad g_i g_i^T = \begin{pmatrix} 0 & \dots & F_i & \dots & 0 \end{pmatrix}$$

$$F = \sum_j g_j g_j^T = G G^T$$

$$\text{Tr}(KF) = \text{Tr}(KG G^T) = \text{Tr}(G^T K G)$$

s.t constraints on G .

(3)

(1) $G \geq 0$

(2) $G^T G_{ij} = 0$ when $i \neq j$ \rightarrow because each point belongs to 1 cluster only
 $G^T G_{ii} = \frac{1}{l_i} \mathbf{1}^T \mathbf{1} = 1$, thus $\boxed{G^T G = I}$

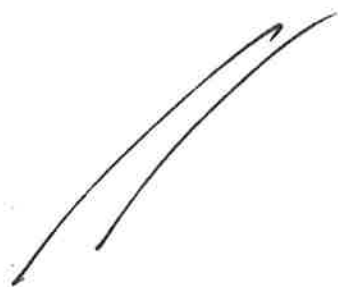
(3) F is doubly stochastic $\Rightarrow G^T G \cdot \vec{1} = \vec{1}$

Therefore, k -means is equivalent to:

$$\max_{G \in \mathbb{R}^{m \times k}} \text{Tr}(G^T K G)$$

(*)

s.t. $\left\{ \begin{array}{l} G \geq 0 \\ G^T G = I \\ G G^T \vec{1} = \vec{1} \end{array} \right.$



The constraint $G^T G = I$ is the req. that each point belongs to 1 cluster only!