k-means

①

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

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

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

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

$$\psi_1, \psi_2, \dots, \psi_k \qquad \psi_i \in \{1, \dots, n\}$$

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

$$\psi_i \cap \psi_j = \phi$$

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

$l_j \to |\psi_j|$, $\quad c_j \to \frac{1}{|\psi_j|} \sum_{i \in \psi_j} x_i$

$$\min - \sum_j \frac{1}{l_j} \sum_{r,s \in \psi_j} x_r^T x_s$$

$$\boxed{\max_{\{\psi_i\}} \sum_{j=1}^{k} \frac{1}{l_j} \sum_{r,s \in \psi_j} x_r^T x_s}$$

$x_i \to \phi(x_i)$

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

$\to k(x_i, x_j)$

"kernel Trick"

kernel Matrix: $K = \left(k(x_i, x_j)\right)_{n \times n}$

symmetric, positive semi-definite

$$F = \left(F_{ij} = \frac{1}{l_r}\right) \text{ if } (i,j) \in \Psi_r$$

$F_{ij} = 0$ otherwise

$$F = \begin{pmatrix} F_1 & & & \\ & F_2 & & \\ & & \ddots & \\ & & & F_n \end{pmatrix}.$$

Block Diagonal

$$F_r = \frac{1}{l_r} \mathbf{1} \cdot \mathbf{1}^T$$

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

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

$$G = \begin{pmatrix} | & & | \\ g_1 & \cdots & g_k \\ | & & | \end{pmatrix} \quad g_r g_r^T = \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & F_r & \\ & & & \ddots \; 0 \end{pmatrix}$$

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

$$Tr(KF) = Tr(KGG^T) = Tr(G^TKG)$$

s.t constraints on G.

(1) $G \geq 0$

(2) $G^TG_{ij} = 0$ when $i \neq j$ → because each point belongs to 1 cluster

$G^TG_{ii} = \frac{1}{l_i} 1^T 1 = 1$, thus $\boxed{G^TG = I}$

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

Therefore, k-means is equivalent to:

$$\max_{G \in \mathbb{R}^{m \times k}} Tr(G^TKG) \quad (*)$$

$$\text{s.t.} \begin{cases} G \geq 0 \\ G^TG = I \\ GG^T\vec{1} = \vec{1} \end{cases}$$

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