Class Sept. 11 2020

Behaviour of J^{clust} w.r.t. different decision variables.

(i) Optimal cluster assignment.

Fix k: no. of clusters. ←

Fix $z_1, z_2, \ldots, z_k$: cluster representatives.

Choose $c_1, c_2, \ldots, c_N$: "best" clustering assignment so that J^{clust} is minimized.

Observe:

$$\min_{c_1,\ldots,c_N} J^{clust} = \min_{c_1,\ldots,c_N} \frac{\|x_1 - z_{c_1}\|^2 + \|x_2 - z_{c_2}\|^2 + \ldots + \|x_N - z_{c_N}\|^2}{N}$$

$$= \frac{1}{N}\left[\min_{c_1} \|x_1 - z_{c_1}\|^2 + \min_{c_2} \|x_2 - z_{c_2}\|^2 + \ldots + \min_{c_N} \|x_N - z_{c_N}\|^2 \right]$$

Find the nearest neighbour amongst $\{z_1, \ldots, z_k\}$ for x_1.

(2) Optimal group representatives.

k: fixed. ←

$G_1, \ldots, G_j$: j clusters.

$c_1, \ldots, c_N$: fixed (assignment is fixed)

$$J^{clust} = J_1 + J_2 + \cdots + J_k$$

where $J_j = \frac{1}{N} \sum_{i \in G_j} \|x_i - z_j\|^2$

$$\min_{z_1, \ldots, z_k} J^{clust} = \min_{z_1} J_1 + \min_{z_2} J_2 + \cdots + \min_{z_k} J_k$$

$$\min_{z_j} J_j = \min_{z_j} \frac{1}{N} \sum_{i \in G_j} \|x_i - z_j\|_2^2$$

$$z_j = \frac{1}{|G_j|} \sum_{i \in G_j} x_i \qquad \text{(H.W.)}$$

$$J^{clust} = \frac{\|x_1 - z_{c_1}\|^2 + \|x_2 - z_{c_2}\|^2 + \cdots + \|x_N - z_{c_N}\|^2}{N}$$

$c_i = 1$

$$= \frac{\left(\|\otimes - z_1\|^2 + \|\otimes - z_1\|^2 + \cdots\right) + \left(\| - z_2\|^2 + \| - z_2\|^2 + \cdots\right) + \cdots + \left(\| - z_k\|^2 + \cdots + \| - z_k\|^2\right)}{N}$$

$$= \frac{(\quad)}{N} + \frac{(\quad)}{N} + \cdots + \frac{(\quad)}{N}$$

$$= J_1 + J_2 + \cdots + J_k$$

J^{clust}

Iterative algo.

k-mean algorithm

Input: $x_1, \ldots, x_N$

initial: k- fixed

$z_1, z_2, \ldots, z_k$: fixed.

Repeat,

1) Partition $x_1, \ldots, x_N$ in k- clusters.

2) Update the representatives