# K-Means Clustering.

$$\sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \longrightarrow \min_{C_1,\dots,C_K}$$

$\Updownarrow$

centroid formula

$$2 \cdot \sum_{k=1}^{K} \sum_{i \in C_k} \sum_{j=1}^{p} (x_{ij} - \bar{x}_{kj})^2$$

$=$

Focus on:

$$(x_{ij} - x_{i'j})^2 = ((x_{ij} - \bar{x}_{kj}) + (\bar{x}_{kj} - x_{i'j}))^2$$

$$= (x_{ij} - \bar{x}_{kj})^2 + 2(x_{ij} - \bar{x}_{kj})(\bar{x}_{kj} - x_{i'j}) + (x_{i'j} - \bar{x}_{kj})^2$$

$$\sum_{i \in C_k} (x_{ij} - x_{i'j})^2 = \sum_{i \in C_k} (x_{ij} - \bar{x}_{kj})^2 + 2 \sum_{i \in C_k} (x_{ij} - \bar{x}_{kj})(\bar{x}_{kj} - x_{i'j}) + \sum_{i \in C_k} (x_{i'j} - \bar{x}_{kj})^2$$

$=0$ !

$\sum_{i' \in C_k}$

$= |C_k| (x_{i'j} - \bar{x}_{kj})^2$