# Exercise 2
27 September 2023 12:00

## Definition of k-means Clustering:

Choose $\hat{m}_1^{km}, \ldots, \hat{m}_k^{km}$ and $c^{km}(1), \ldots, c^{km}(n)$ so that

$$\boxed{S(C, m_1, \ldots, m_k) = \sum_{i=1}^{n} \|x_i - m_{c(i)}\|^2}$$

is minimal

**Proof:**

Let $\tilde{x}_i = q x_i \quad \forall i \in \mathbb{N}_n$

We know for $\|\cdot\|$ that $\|qx\| = |q|\,\|x\|$, if $q \in \mathbb{R}$

Let also $\tilde{m}_{\tilde{c}(i)}$ be the optimal centroids for $\tilde{x}_i$

$$\Rightarrow \tilde{S}(C, m_1, \ldots, m_k) = \sum_{i=1}^{n} \|\tilde{x}_i - \tilde{m}_{\tilde{c}(i)}\| = \sum_{i=1}^{n} \|q(x_i - q^{-1}\tilde{m}_{\tilde{c}(i)})\| = \overbrace{|q|}^{>0} \sum_{i=1}^{n} \|x_i - q^{-1}\tilde{m}_{\tilde{c}(i)}\|$$

$\Rightarrow$ Minimal exactly when $\sum_{i=1}^{n} \|x_i - y_{t(i)}\|$ is minimal with $y_{t(i)} = q^{-1}\tilde{m}_{\tilde{c}(i)}$, $t : \mathbb{N}_n \to \mathbb{N}_k$

But we also know from the definition of the k-means clustering:

$\sum_{i=1}^{n} \|x_i - y_{t(i)}\|$ is minimal $\Leftrightarrow$ $y_{t(i)} = m_{c(i)}$

$\Rightarrow m_{c(i)} = q^{-1}\tilde{m}_{\tilde{c}(i)} \Leftrightarrow q m_{c(i)} = \tilde{m}_{\tilde{c}(i)}$

$\Rightarrow q m = \tilde{m}$ and $c = \tilde{c}$ except for permutations/order of the clusters