CLUSTERING

MAIN PROBLEM :

$$\begin{cases} \min \sum_{i=1}^{\ell} \min_{j=1,\dots,K} d(p_i, x_j) \\ x_j \in \mathbb{R}^n \quad \forall j = 1,\dots,K \end{cases}$$

$x_j :=$ CENTROID

$\|\cdot\|_2$

$$\begin{cases} \min \sum_{i=1}^{\ell} \min_{j=1,\dots,K} \|p_i - x_j\|_2^2 \\ x_j \in \mathbb{R}^n \quad \forall j = 1,\dots,K \end{cases}$$

TAKEN $\alpha_{ij}^* = \begin{cases} 1 & \text{IF } \|p_i - x_j\|_2 = \min_{h=1,\dots,K} \|p_i - x_j\|_2 \\ 0 & \text{OTHERWISE} \end{cases}$

$$\begin{cases} \min_{x,\alpha} f(x,\alpha) := \sum_{i=1}^{\ell} \sum_{j=1}^{K} \alpha_{ij} \|p_i - x_j\|_2^2 \\ \sum_{i=1}^{K} \alpha_{ij} = 1 \quad \forall i = 1,\dots,\ell \\ \alpha_{ij} \geq 0 \\ x_j \in \mathbb{R}^n \end{cases}$$

THE **K-MEANS** ALGORITHM CONSISTS IN AN ALTERNATING MINIMIZATION

1) FIXED x_j WE SOLVE ℓ LP PROBLEMS AND WE ASSIGN EACH POINT TO THE BEST CLUSTER.

$\forall i = 1,\dots,\ell \Rightarrow \alpha_{ij}^*$

2) FIXED α_{ij} WE SOLVE K CONVEX QUADRATIC PROBLEMS

$\Rightarrow$ UPDATE THE CENTROIDS $\Rightarrow \forall j = 1,\dots,K \quad x_j' = \dfrac{\sum_{i=1}^{\ell} \alpha_{ij} p_i}{\sum_{i=1}^{\ell} \alpha_{ij}}$ (MEAN)

$\|\cdot\|_1$

$$\begin{cases} \min \sum_{i=1}^{\ell} \min_{j=1,\dots,K} \|p_i - x_j\|_1 \\ x_j \in \mathbb{R}^n \quad \forall j = 1,\dots,K \end{cases}$$

$$\begin{cases} \min_{x,\alpha} f(x,\alpha) := \sum_{i=1}^{\ell} \sum_{j=1}^{K} \alpha_{ij} \|p_i - x_j\|_1 \\ \sum_{i=1}^{K} \alpha_{ij} = 1 \quad \forall i = 1,\dots,\ell \\ \alpha_{ij} \geq 0 \\ x_j \in \mathbb{R}^n \end{cases}$$

K-MEDIAN

1)

2) FIXED α_{ij} WE SOLVE K SIMPLE CONVEX PROBLEMS

$\Rightarrow \forall j = 1,\dots,K \Rightarrow x'_j = \text{median}(p_i : \alpha_{ij} = 1)$