Clustering Method

- k-means clustering
- Gaussian mixture

1. k-means clustering

$D = \{x_i\}_{i=1}^{N}$ in R^d, positive integer k, partition D into k groups.

where each data point belongs to one (and only one) cluster

One-hot vector $r_i \in \{0,1\}^k$, $i = 1, 2, \cdots, N$, $(r_i)_j = \begin{cases} 1, & \text{if } x_i \text{ is assigned to cluster } j \\ 0, & \text{otherwise} \end{cases}$

Each cluster has a "representative", $z_j \in R^d$, $j = 1, 2, \cdots, k$

Determine $\{z_j\}_{j=1}^{k}$ and $\{r_i\}_{i=1}^{N}$

The objective function is the sum of the squared distance from each data point to the center of the cluster that it is assigned to.

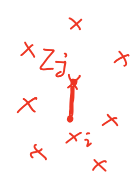
$$\Rightarrow \textbf{Min} \quad J(\{r_i\}, \{z_j\}) = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{1}{2} (r_i)_j \, |x_i - z_j|^2$$

(Note : $(r_i)_j = 1 \iff x_i \in$ cluster j)

Step 1 : Assignment step

Given $\{z_j\}_{j=1}^{k}$, assign each x_i to the nearest z_j

$$(r_i)_j = \begin{cases} 1, & \text{if } j = \arg\min_k |x_i - z_k|^2 \\ 0, & \text{otherwise} \end{cases}$$



Step 2 : Update step

Given $\{r_i\}_{i=1}^{N}$, set $z_j = \dfrac{\sum_{i=1}^{N} (r_i)_j x_i \;①}{\sum_{i=1}^{N} (r_i)_j \;②}$, $j = 1, 2, \cdots, k$.

$$\left(0 = \nabla_{z_j} J = -\frac{1}{N} \sum_{i=1}^{N} (r_i)_j (\underline{x_i} - \underline{z_j}) \Rightarrow 0 = \sum_{i=1}^{N} (r_i)_j \underline{x_i} - \left(\sum_{i=1}^{N} (\underline{r_i})_j \right) z_j \right)$$

z_j: average (mean) of x_i's that have been assigned to cluster j.

① : sum of samples in j-th cluster

② : number of data point in j-th cluster

Algorithm (K-means)

- Input data $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^N$
- Specify the number of cluster K
- Initialize the cluster centers $\{\mathbf{z}_j\}_{j=1}^K$
- Repeat:

  update $\{\mathbf{r}_i\}_{i=1}^N$ (step 1)

  update $\{\mathbf{z}_j\}_{j=1}^K$ (step 2)

  until convergence
- Return the cluster centers $\{\mathbf{z}_j\}_{j=1}^K$ and assignment $\{\mathbf{r}_i\}_{i=1}^N$

Convergence:

- For a finite dataset, the algorithm converges in a finite number of iterations.
- The algorithm may converge to a local minimum of the objective function.

Example:

$K=3$