K-Means Clustering:

Non-parametric method for clustering ⇒ Operate by considering the # of clusters unknown!

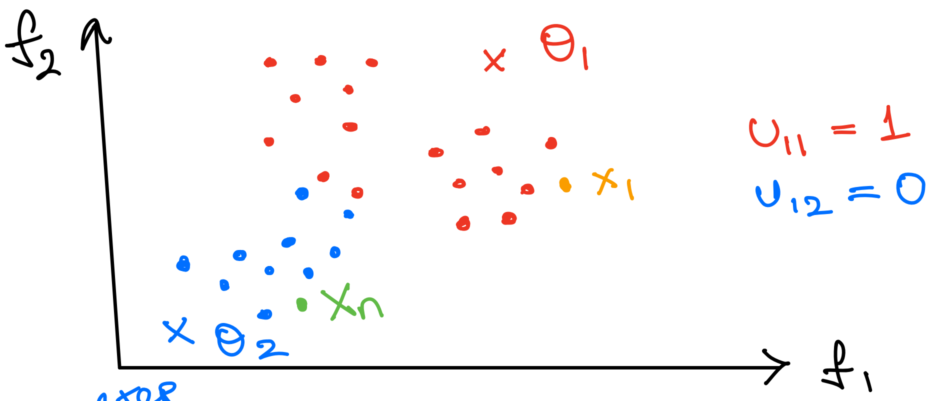
parametric clustering ⇒ # of clusters known a priori!

k = # of groups or clusters the algorithm will find in the data.

① Centroid-based clustering

FOLLOWS EM ALGORITHM

Step 1: Initialize the cluster centroids for each cluster, Θ_j, where $j \in \{1, \ldots\ldots, k\}$



	1	2
1	1	0
2	⋮	
⋮		
n	0	1

E-step

Step 2: Assignment membership to each sample, u_{ik}

NOTE: HARD MEMBERSHIP ASSIGNMENT!

(Implies each data point can only belong to one cluster)

$u_{ik} \equiv$ membership of sample x_i in cluster with centroid k

$u_{ik} \in \{0, 1\}$ such that

$$\sum_{j=1}^{K} u_{ij} = 1 \quad \forall i$$

Cluster assignment is based on which cluster centroid the sample x_i is closest to!

M-Step

Step 3: Fix memberships and update cluster centroids,

$$\theta_j = \underbrace{\sum_{i \in \theta_j} \frac{x_i}{N_j}}$$

sample average of points assigned to θ_j

Step 4: go back to step 2 and iterate until convergence criteria is met.

No significant change in position of cluster centroid.

COMPUTING MEMBERSHIPS:

$U \equiv$ membership matrix, $\underbrace{N}_{\#\text{ samples}} \times \overbrace{K}^{\#\text{ clusters}}$

$$U = \begin{array}{c|cccc} & 1 & 2 & \cdots & K \\ 1 & 0 & 1 & \cdots & 0 \\ 2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ N & 0 & 0 & \cdots & 1 \end{array}$$

$\rightarrow K$ $\qquad \sum_{j=1} u_{ij} = 1$

$\rightarrow$ Predictions

$\rightarrow \sum_{i=1}^{N} u_{i2} = N_2 \equiv$ # samples assigned to cluster 2.

Label assignment $\{1, 2, \ldots, K\}$

$u_{ij} \equiv$ membership of point x_i to cluster with centroid θ_j, $u_{ij} \in \{0, 1\}$

$\theta_j \equiv$ cluster centroid for cluster j

$x_i \equiv$ i^{th} input sample

$d(x_i, \theta_j) \equiv$ distance between x_i and θ_j

$$\boxed{J(U, \theta) = \sum_{i=1}^{N} \sum_{j=1}^{K} u_{ij} \, d^2(x_i, \theta_j) \quad \text{such that} \quad \sum_{j=1}^{K} u_{ij} = 1, \ \forall i}$$

K-Means objective function

$K \to N \Rightarrow J \to 0$, every point becomes its own cluster centroid. Trivial solution that we are not interested in. $\underline{\underline{K}}$ can be selected with cluster validity metrics.

Distance metrics:

① Euclidean

$$d_{Euclidean}(x_i, \theta_j) = d_E(x_i, \theta_j)$$
$$= \underbrace{\| x_i - \theta_j \|_2}_{L2\text{-norm}}$$

Lp_norm

$\{p \geq 1\} \in \mathbb{R}$

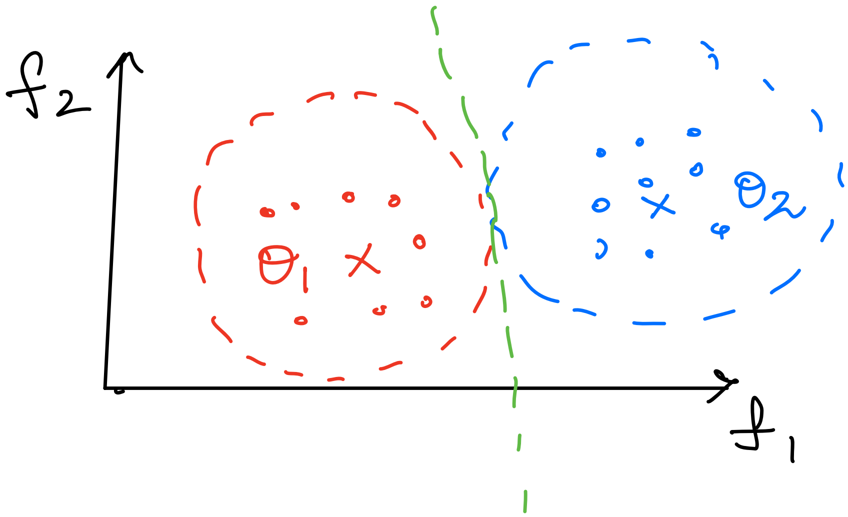
vector $x = (x_1, \ldots, x_n)$

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Infinity or maximum norm $\left\{ \|x\|_\infty := \max_i |x_i| \right.$

$$J(\theta, U) = \sum_{i=1}^{N} \sum_{j=1}^{K} u_{ij} \| x_i - \theta_j \|_2^2$$

$$\text{s.t.} \sum_{j=1}^{K} u_{ij} = 1, \quad \forall i$$



cluster shapes are circular or spherical-like with Euclidean distance.

(2) Mahalanobis Distance:

$$d_M^2(x_i, \theta_j) = (x_i - \theta_j)^T \Sigma_j^{-1} (x_i - \theta_j)$$

NOTE: need to also estimate the value of covariance matrix (Σ) for each cluster group, Σ_j.

If Σ_j is singular:

(1) remove linearly dependent features
(2) diagonally load Σ_j.

(3) Manhattan distance :

$$d_{Man}(x_i, \theta_j) = |x_i - \theta_j|$$
$$= \|x_i - \theta_j\|_1$$

(4) Cosine similarity :
measures similarity between two non-zero vectors.

$$\text{Cosine similarity}(a, b) = \frac{a \cdot b}{\|a\|_2 \|b\|_2}$$

$$d_{cosine}(a, b) = 1 - \frac{a \cdot b}{\|a\|_2 \|b\|_2}$$

- useful to calculate distance between two sparse vectors.