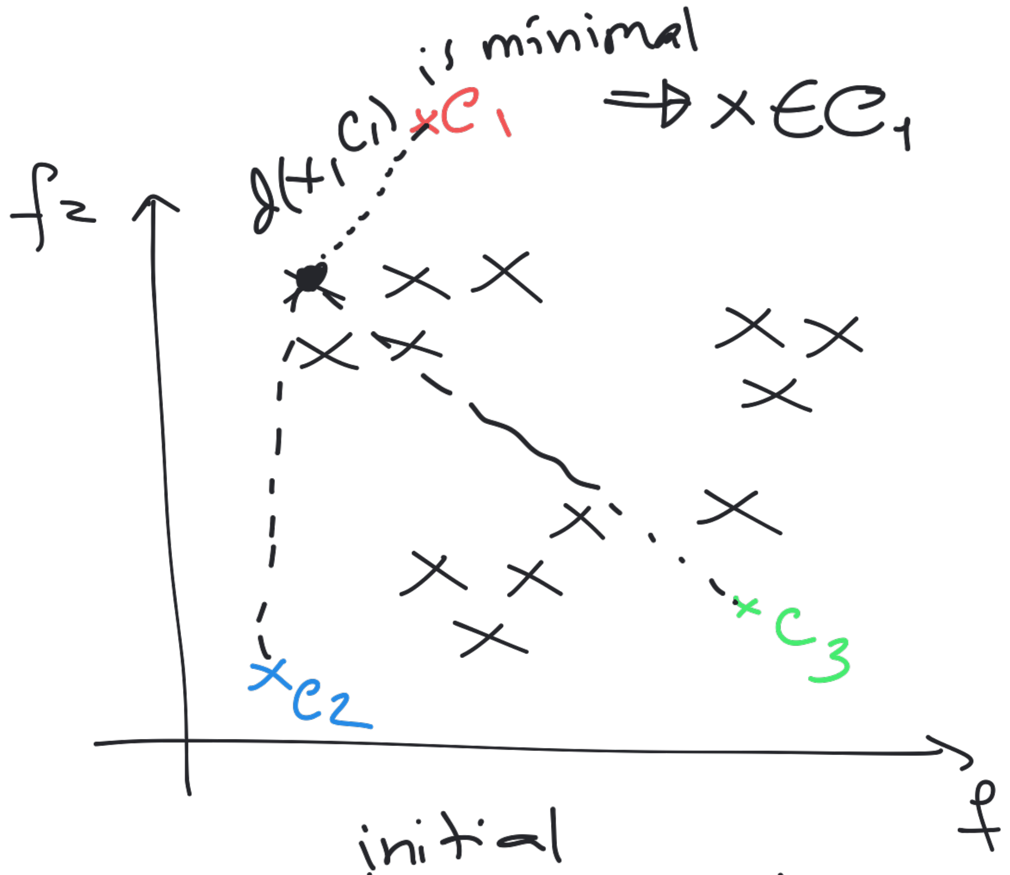
Clustering

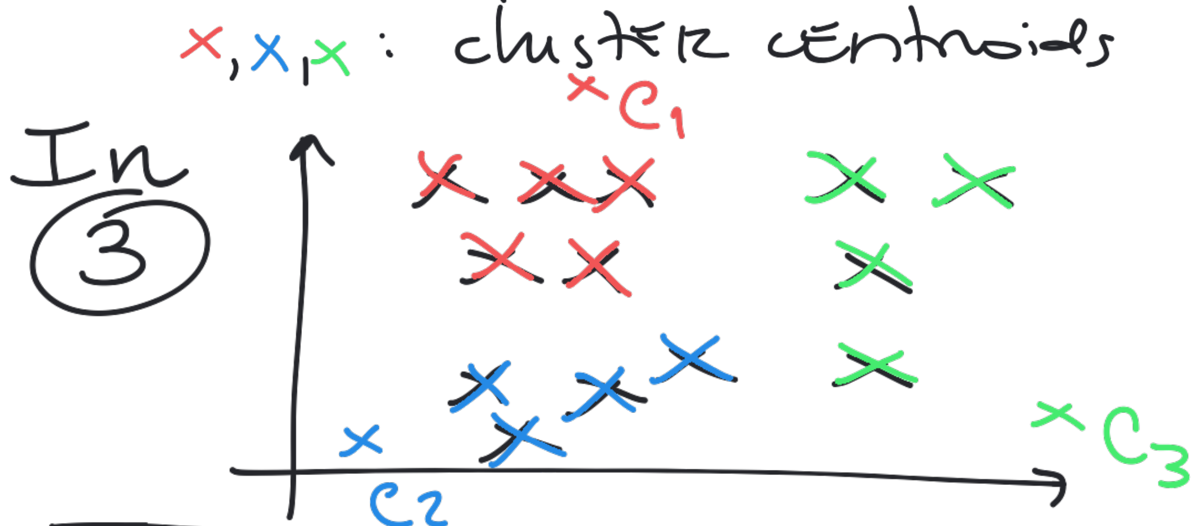
GMM

↳ parametric characterization of the data

K-Means Algorithm

↳ non-parametric charecterization of the data





PSEUDO-CODE FOR K-MEANS

① Specify # of clusters $\underline{\underline{K}}$.

② Initialize the cluster representatives or centroids of the clusters

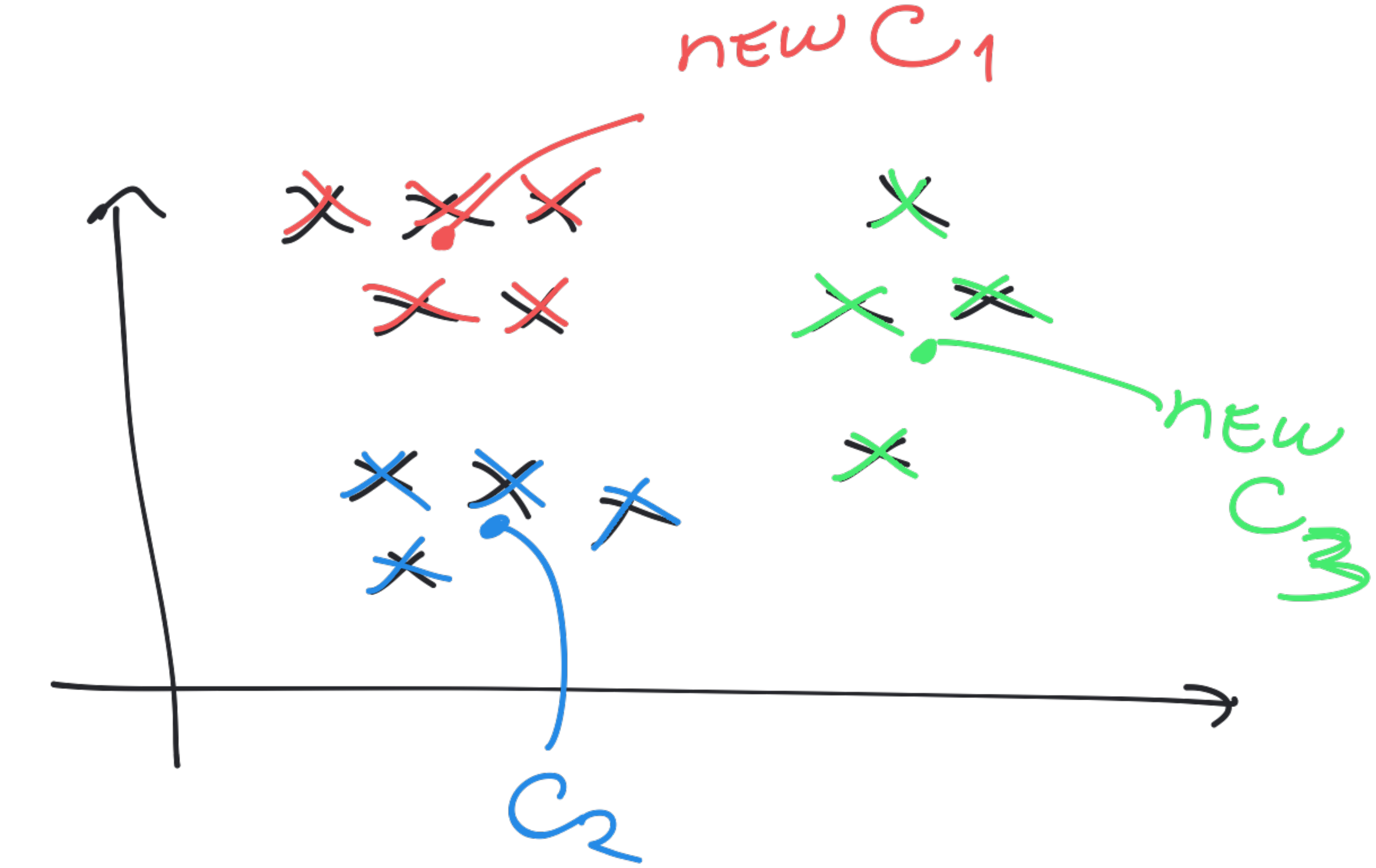
③ Determine closest centroid to each point.

$$D_{N\times K} = \begin{array}{c} \\ x_1 \\ x_2 \\ \vdots \\ x_N \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & K \end{array} \\ \begin{bmatrix} \bigcirc & & & \\ & & & \\ & & & \\ & & & \end{bmatrix} \end{array}$$

$d(x_1, c_1)$

④ Update the centroids C_K using the mean of all points assigned to C_K.

⑤ Repeat steps ② and ④ until convergence



In 4

Alternating optimization:

1. first label each sample for a given (and fixed) set of cluster centroids.
2. Using this labeling, update the cluster centroids using the average of points assign to cluster

Disadvantages of E.M. or alternating optimization:

1. Depending on initialization, we will converge to $\neq$ solutions.

→ May get converge to LOCAL MINIMA.

Objective function of K-Means:

$$J(\theta, U) = \sum_{i=1}^{N} \sum_{k=1}^{K} u_{ik} \cdot d_E^2(x_i, \theta_k)$$

cluster centroids

$\theta = \{\theta_k\}_{k=1}^{K}$

Memberships for each point

$$\boxed{\arg\min_{\{\theta, U\}} J(\theta, U)}$$

$u_{ik} = \{0, 1\}$

MEMBERSHIP MATRIX

In GMM:

$$C_{ik} = p(z_i = k \mid x, \theta^t)$$

$\equiv$ prob. of point x_i in cluster k

Soft assignment

In K-Means:

$$U_{ik} = \{0, 1\}$$

$\equiv$ membership of point x_i in cluster k

Crisp assignment

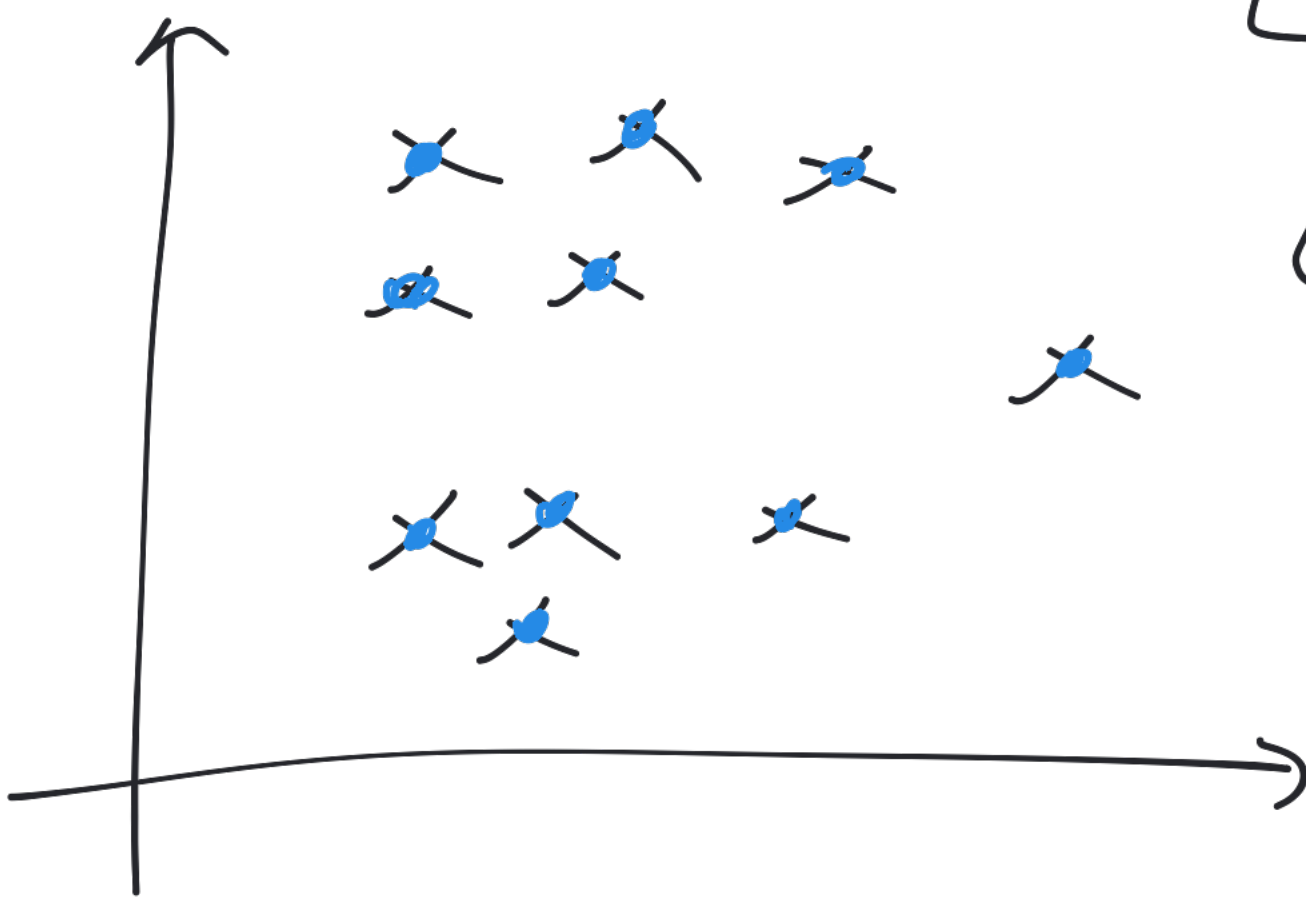
OR

Hard

→ sometimes K-means is used as initialization technique for GMM.

$$J(\theta, U) = \sum_{i=1}^{N} \sum_{k=1}^{K} \underline{\underline{u_{ik}}}\, d^2(x_i, \underline{\underline{\theta_k}})$$

- what happens when $\underline{\underline{K = N}}$?



$\hookrightarrow$ $J = 0$

$\hookrightarrow$ we cannot use obj. fct. J to find # clusters K.

Ⓐ $$\frac{\partial J}{\partial u_{ik}} = \sum_{i=1}^{N} d^2(x_i, \theta_k) = 0$$

$$\Longrightarrow u_{ik} = \begin{cases} 1 & \text{if } k = \arg\min d^2(x_i, \theta_k) \\ 0 & \text{otherwise} \end{cases}$$

Ⓑ $$\frac{\partial J}{\partial \theta_k} = 0$$

to be continued...