CLUSTERING

$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}$ → class labels (y_i)

↳ data vectors (x_i)

Binary $\Rightarrow y_i \in \{0,1\}$ or $\{-1,+1\}$

Multiclass $\Rightarrow y_i \in \{1, 2, \ldots, K\}$

$\mathcal{X} = \{x_i\}_{i=1}^{N} \Rightarrow$ no class labels.

We assumed a certain density form for classes.

$p(x \mid y=c) \rightarrow$ estimated the parameters.

Parametric classification

$\hat{\Sigma}_1, \hat{\mu}_1$, $\hat{p}(y=1)$

$\underline{p(x \mid y=1)}$ $\qquad$ $\underline{p(x \mid y=2)}$ $\qquad$ $\underline{p(x \mid y=3)}$

$\underline{p(y=1)}$ $\qquad$ $\underline{p(y=2)}$ $\qquad$ $\underline{p(y=3)}$

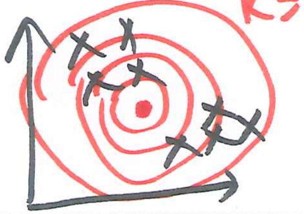
$\underline{p(y \mid x) = ?}$

Mixture Densities

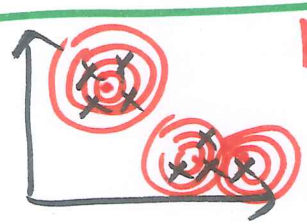
$$\boxed{p(x) = \sum_{k=1}^{K} \underbrace{p(x \mid C_k)}_{\text{component density}} \underbrace{p(C_k)}_{\text{mixture proportion}}}$$

C_k : cluster k

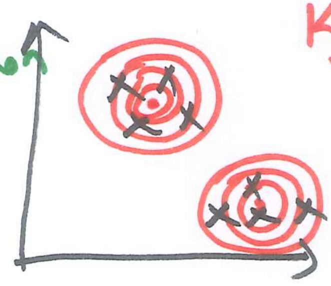
K=1



K=3



K=2



$\underline{\underline{\Phi}} = \{ \hat{p}(c_k), \hat{\mu}_k, \hat{\Sigma}_k \}_{k=1}^{K}$ K: # of components.

$y_{ik} = \begin{cases} 1 & \text{if } x_i \text{ belongs to component } k \text{ cluster.} \\ 0 & \text{otherwise} \end{cases}$ } one-of-K encoding.

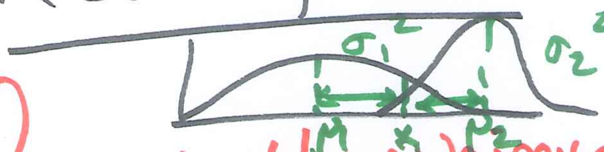
y_{ik} → cluster/component membership.

WE DO NOT KNOW "y_{ik}" VALUES !!!

iterative algorithm ⇒ STEP 1 Estimate the labels.
STEP 2 Estimate the parameters.

$$\hat{p}(c_k) = \frac{\sum_{i=1}^{N} y_{ik}}{N} \qquad \hat{\mu}_k = \frac{\sum_{i=1}^{N} y_{ik} \cdot x_i}{\sum_{i=1}^{N} y_{ik}} \qquad \hat{\Sigma}_k = \frac{\sum_{i=1}^{N} y_{ik}(x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T}{\sum_{i=1}^{N} y_{ik}}$$

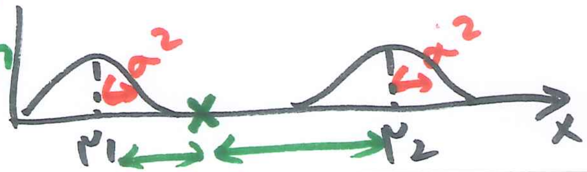
K-MEANS ALGORITHM: K clusters ⇒ K center points.

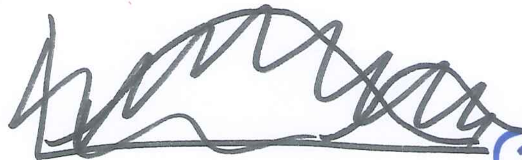


$$\arg\min_k \|x_i - \hat{\mu}_k\|_2$$

$$\left.\begin{array}{l} \|x_i - \hat{\mu}_1\|_2 \\ \|x_i - \hat{\mu}_2\|_2 \\ \vdots \\ \|x_i - \hat{\mu}_K\|_2 \end{array}\right\} \text{pick the minimum.}$$

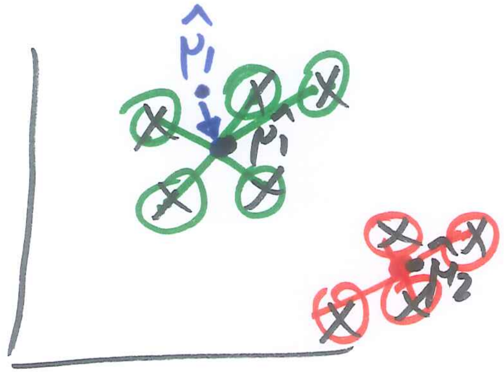
$\exp\left(-\frac{(x-\mu)^2}{\cancel{2\sigma^2}}\right) \cancel{\frac{1}{\sqrt{2\pi\sigma^2}}}$





$$\text{Error} = \sum_{i=1}^{N} \sum_{k=1}^{K} b_{ik} \| x_i - \hat{\mu}_k \|_2^2 \qquad b_{ik} = \begin{cases} 1 & \text{if } \|x_i - \hat{\mu}_k\| = \min_c \|x_i - \hat{\mu}_c\| \\ 0 & \text{otherwise} \end{cases}$$

$$\boxed{\begin{array}{ll} \text{minimize:} & \text{Error} \\ \text{with respect to:} & \underbrace{\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_K}_{\text{SET 1}}, \underbrace{\{b_{ik}\}_{i=1,k=1}^{N,K}}_{\text{SET 2}} \end{array}}$$



Initialize $\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_K$ randomly

Repeat

E-step $\left[\begin{array}{l} \text{for all } x_i \quad i = 1, \ldots, N \\ b_{ik} = \begin{cases} 1 & \text{if } \|x_i - \hat{\mu}_k\| = \min_c \|x_i - \hat{\mu}_c\| \\ 0 & \text{otherwise} \end{cases} \end{array}\right.$

M-step $\left[\begin{array}{l} \text{for all } \hat{\mu}_k . \quad k = 1, \ldots K \\ \hat{\mu}_k = \dfrac{\sum_{i=1}^{N} b_{ik} x_i}{\sum_{i=1}^{N} b_{ik}} \end{array}\right.$

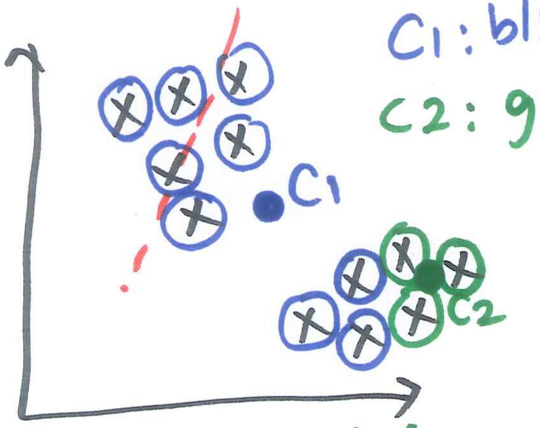
Until convergence. [[all b_{ik}'s stay same]] or

[[all $\hat{\mu}_c$'s stay same]]

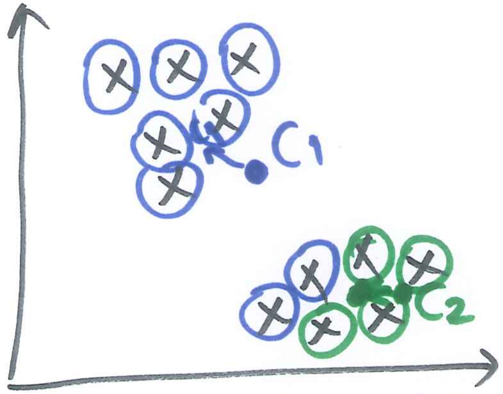
③

C_1 : blue

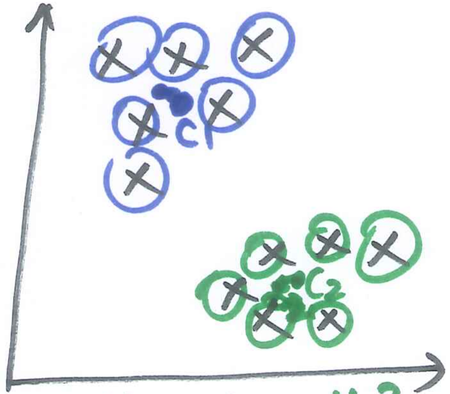
C_2 : green



iteration #1



iteration #2



iteration #3

Expectation - Maximization Algorithm (EM)

$\mathcal{X} = \{x_i\}_{i=1}^{N}$

Likelihood $\ni \mathcal{L}(\Phi \mid \mathcal{X}) = \log \prod_{i=1}^{N} p(x_i \mid \Phi)$

$$= \sum_{i=1}^{N} \log \underbrace{\left[\sum_{k=1}^{K} p(x_i \mid C_k) \cdot p(C_k) \right]}_{\text{mixture densities.}}$$

two sets of random variables

$\rightarrow Z$: cluster memberships. $\rightarrow$ hidden variables. $\underline{\underline{N \times K}}$

$\rightarrow \Phi$: parameters. $\mu_1, \Sigma_1, \mu_2, \Sigma_2, \ldots \mu_K, \Sigma_K$

$\mathcal{L}(\Phi \mid \mathcal{X}, Z)$

EM ALGORITHM:

E-STEP: $E\left[L_c(\Phi \mid \mathcal{X}, \mathcal{Z}) \mid \mathcal{X}, \Phi^{(t)}\right]$

(t) → iteration index.

M-STEP: $\Phi^{(t+1)} = \arg\max \; E\left[L_c(\Phi \mid \mathcal{X}, \mathcal{Z}) \mid \mathcal{X}, \Phi^{(t)}\right]$

$x_i \Rightarrow h_{i1} = 0.8 \quad h_{i2} = 0.2 \quad h_{i3} = 0.0$

E-Step: $h_{ik} = E\left[z_{ik} \mid \mathcal{X}, \Phi^{(t)}\right] = \dfrac{\boxed{p(x_i \mid C_k, \Phi^{(t)})}\, P(C_k)}{\sum_{c=1}^{K} p(x_i \mid C_c, \Phi^{(t)})\, P(C_c)}$

h_{ik} ⇓ success probability.

$h_{ik} \geq 0, \quad \sum_{k=1}^{K} h_{ik} = 1, \; \forall i$

M-Step: $\hat{\mu}_k^{(t+1)} = \dfrac{\sum_{i=1}^{N} h_{ik} \cdot x_i}{\sum_{i=1}^{N} h_{ik}}$

$\hat{\Sigma}_k^{(t+1)} = \dfrac{\sum_{i=1}^{N} h_{ik} (x_i - \hat{\mu}_k^{(t+1)})(x_i - \hat{\mu}_k^{(t+1)})^T}{\sum_{i=1}^{N} h_{ik}}$

multivariate gaussian

multivariate gaussian.

$\hat{p}(C_1) + \hat{p}(C_2) + \ldots + \hat{p}(C_K) = 1$

$\hat{p}(C_k) = \dfrac{\sum_{i=1}^{N} h_{ik}}{N}$

$\dfrac{\sum_{i=1}^{N} h_{i1} + \sum_{i=1}^{N} h_{i2} + \ldots + \sum_{i=1}^{K} h_{ik}}{N} = \dfrac{N}{N} = 1$