Clustering

Classification $\left[\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N} \right.$

$y_i \to$ class labels

$x_i \to$ data points

Binary classification

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

Multiclass classification

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

clustering $\left[\mathcal{X} = \{x_i\}_{i=1}^{N} \right.$ NO CLASS LABELS!

PARAMETRIC CLASSIFICATION

- We assumed that each class follows a certain density
  $p(x \mid y=c)$
- We estimated the parameters

$$p(x \mid y=1) \quad P(y=1) \quad \cdots\cdots \quad p(x \mid y=K) \quad P(y=K)$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad\qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$\hat{\mu}_1, \hat{\Sigma}_1 \qquad \hat{P}(y=1) \qquad P(y=c \mid x) = ? \qquad \hat{\mu}_K, \hat{\Sigma}_K \qquad \hat{P}(y=K)$$

Mixture Densities

K different clusters (unknown)

C_k = cluster # k.

$$p(x) = \sum_{k=1}^{K} \underbrace{p(x|C_k)}_{\text{component density}} \cdot \underbrace{P(C_k)}_{\text{mixture proportions}}$$

$$\Phi = \{ \hat{P}(C_k), \hat{\mu}_k, \hat{\Sigma}_k \}_{i=1}^{K}$$

K = # of components

$$y_{ik} = \begin{cases} 1 & \text{if } x_i \text{ belongs to component/cluster } k \\ 0 & \text{otherwise} \end{cases}$$

↳ cluster/component membership

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

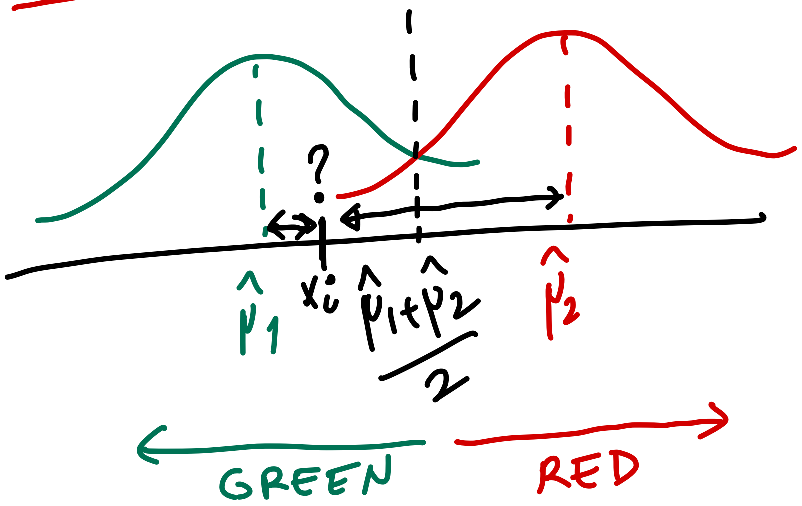
Iterative algorithm:

STEP 1: Estimate the cluster memberships.

STEP 2: Estimate the parameters.

$$\hat{P}(c_k) = \frac{\sum_{i=1}^{N} \hat{y}_{ik}}{N} \qquad \hat{\mu}_k = \underbrace{\frac{\sum_{i=1}^{N} \hat{y}_{ik}\, x_i}{\sum_{i=1}^{N} \hat{y}_{ik}}} \qquad \hat{\Sigma}_k = \frac{\sum_{i=1}^{N} \hat{y}_{ik}(x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T}{\sum_{i=1}^{N} \hat{y}_{ik}}$$

K-MEANS CLUSTERING



$$\exp\left[-\frac{(x_i - \hat{\mu}_1)^2}{\cancel{2\hat{\sigma}_1^2}}\right] \cdot \cancel{\frac{1}{\sqrt{2\pi\hat{\sigma}_1^2}}}$$

$$\exp\left[-\frac{(x_i - \hat{\mu}_2)^2}{\cancel{2\hat{\sigma}_2^2}}\right] \cdot \cancel{\frac{1}{\sqrt{2\pi\hat{\sigma}_2^2}}}$$

$$\hat{\sigma}_1^2 = \hat{\sigma}_2^2$$

$$P(y=1 \mid x) = \frac{p(x \mid y=1)\, P(y=1)}{p(x)}$$

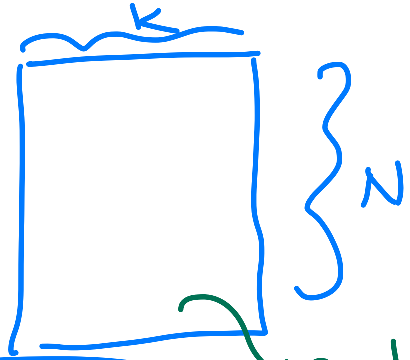
$$P(y=2 \mid x) = \frac{p(x \mid y=2)\, P(y=2)}{p(x)}$$

$$\underbrace{\|x_i - \hat{\mu}_1\|_2 \quad \|x_i - \hat{\mu}_2\|_2 \quad \cdots \quad \|x_i - \hat{\mu}_k\|_2}$$

assume 2nd distance is minimum

$$\hat{y}_{i1} = 0 \quad \hat{y}_{i2} = 1 \quad \hat{y}_{i3} = 0 \quad \cdots \quad \hat{y}_{ik} = 0$$

$$\text{Error} = \sum_{i=1}^{N}\sum_{k=1}^{K} b_{ik}\,\|x_i - \hat{\mu}_k\|_2^2$$

$B=$ 

each row contains exactly one 1.

$$b_{ik} = \begin{cases} 1 & \text{if } \|x_i - \hat{\mu}_k\|_2 = \min_{c=1}^{K} \|x_i - \hat{\mu}_c\|_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\boxed{\begin{array}{l} \text{minimize} \quad \sum_{i=1}^{N}\sum_{k=1}^{K} b_{ik}\,\|x_i - \hat{\mu}_k\|_2^2 \\ \text{with respect to: } \hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_K,\ \{b_{ik}\}_{i=1,k=1}^{N,K} \end{array}}$$

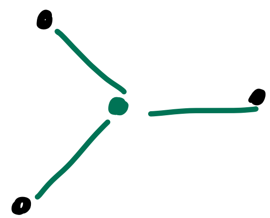
- Initialize $\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_K$ randomly
- Repeat

E-STEP
for all x_i

$$b_{ik} = \begin{cases} 1 & \text{if } \|x_i - \hat{\mu}_k\|_2 = \min_{c=1}^{K} \|x_i - \hat{\mu}_c\|_2 \\ 0 & \text{otherwise} \end{cases}$$

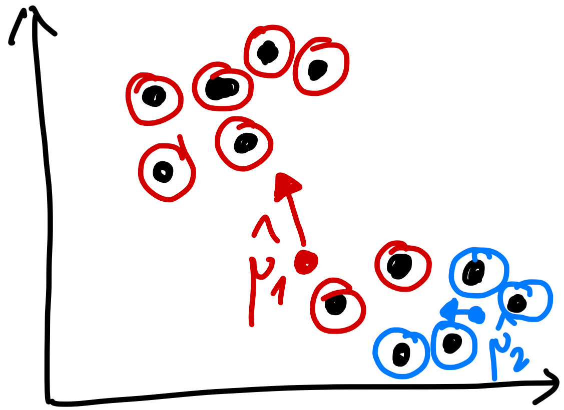
M-STEP
for all $\hat{\mu}_k$

$$\hat{\mu}_k = \frac{\sum_{i=1}^{N} b_{ik} \cdot x_i}{\sum_{i=1}^{N} b_{ik}}$$

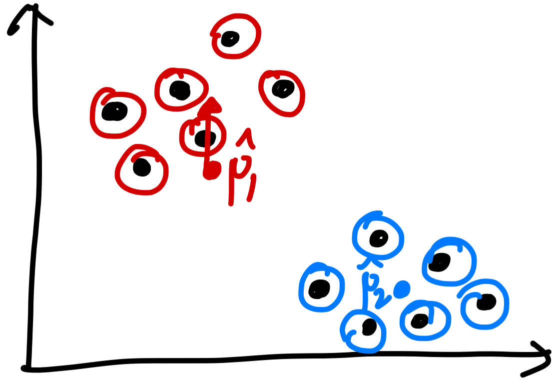


- Until convergence [all b_{ik}'s stay the same] or [all $\hat{\mu}_k$'s stay the same]

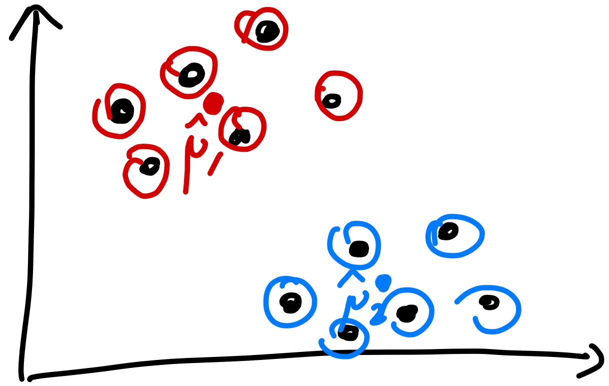
$K=2$



Iteration #1

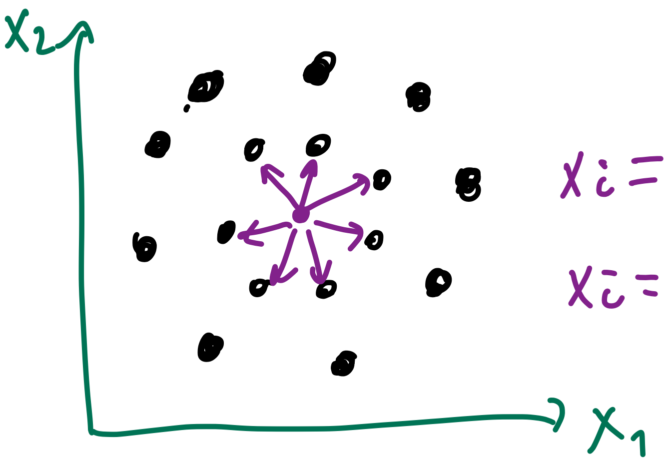


Iteration #2



Iteration #3

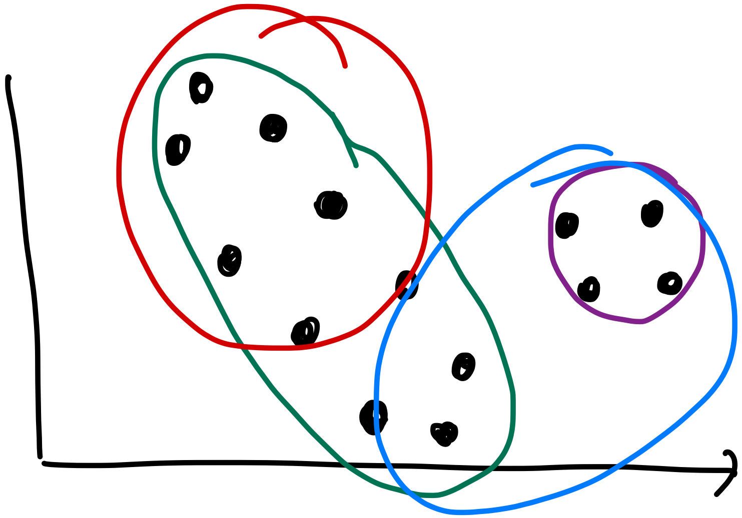
b_{ik}'s stay the same.



$x_i = (d_i, \theta_i)$

$x_i = (x_{i1}, x_{i2})$

$z_i = \Phi(x_i)$



assumes different covariances

assumes shared covariance

Expectation - Maximization (EM) Algorithm

$\mathcal{X} = \{x_i\}_{i=1}^{N}$ loglikelihood $\Rightarrow L(\Phi \mid \mathcal{X}) = \log \left[\prod_{i=1}^{N} p(x_i \mid \Phi) \right]$

$$\mathcal{L}(\Phi \mid \mathcal{X}) = \sum_{i=1}^{N} \log \Big[\underbrace{\sum_{k=1}^{K} p(x_i \mid C_k) P(C_k)}_{\text{mixture densities}} \Big]$$

two sets of random variables

Z = cluster memberships (hidden variables)

Φ = parameters $[\hat{\mu}_1, \hat{\mu}_2, \dots \hat{\mu}_K, \hat{\Sigma}_1, \hat{\Sigma}_2, \dots \hat{\Sigma}_K]$

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

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

multivariate Gaussian

E-STEP:

$$h_{ik} = E\left[z_{ik} \mid x, \Phi^{(t)}\right] = \frac{p(x_i \mid c_k, \Phi^{(t)}) \cdot P(c_k)}{\sum_{c=1}^{K} p(x_i \mid c_c, \Phi^{(t)}) \cdot P(c_c)}$$

$y_i \Rightarrow [0 \quad 1 \quad 0]$

$\hat{y}_i \Rightarrow [0.2 \quad 0.7 \quad 0.1]$

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

M-STEP:

$$\hat{\mu}_k^{(t+1)} = \frac{\sum_{i=1}^{N} h_{ik} \cdot x_i}{\sum_{i=1}^{N} h_{ik}}$$

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

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