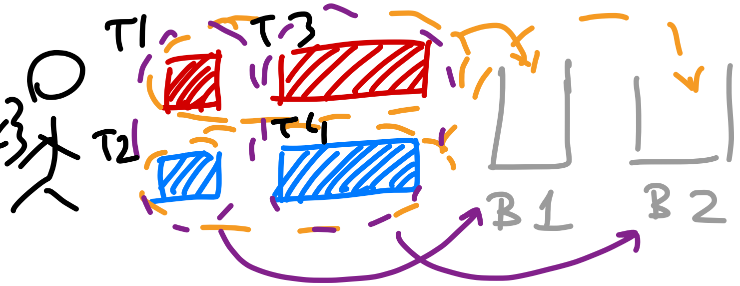
Clustering

classification data set $\left[\ \mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}\right.$ $\xrightarrow{}$ class labels

$y_i \in \{0,1\}$

$y_i \in \{-1,+1\}$ / $y_i \in \{1,2,..k\}$



clustering data set $\left[\ \mathcal{X} = \{x_i\}_{i=1}^{N}\right.$ $\Rightarrow$ NO CLASS LABELS! $X = \begin{array}{l} \begin{bmatrix} \cancel{\text{Red}} & \text{Small} \\ \cancel{\text{Blue}} & \text{Small} \\ \cancel{\text{Red}} & \text{Large} \\ \cancel{\text{Blue}} & \text{Large} \end{bmatrix} \end{array} \begin{array}{l} T1 \\ T2 \\ T3 \\ T4 \end{array}$

PARAMETRIC CLASSIFICATION

- We assumed that each class follows a certain density

$$p(x \mid y=c)$$

- We estmated the parameters

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

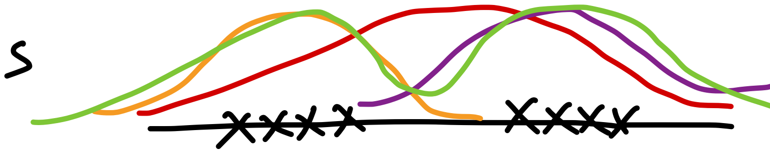
$$\downarrow \qquad\quad \downarrow \qquad\qquad\qquad \downarrow \qquad\quad \downarrow$$

$$\hat{\mu}_1, \hat{\Sigma}_1 \quad \hat{\Pr}(y=1) \quad \cdots \quad \hat{\mu}_K, \hat{\Sigma}_K \quad \hat{P}_r(y=K)$$

$$\Pr(y=c \mid x) = ?$$

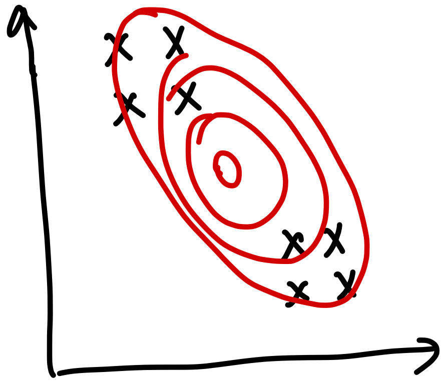
Mixture Densities

K different clusters (unknown)

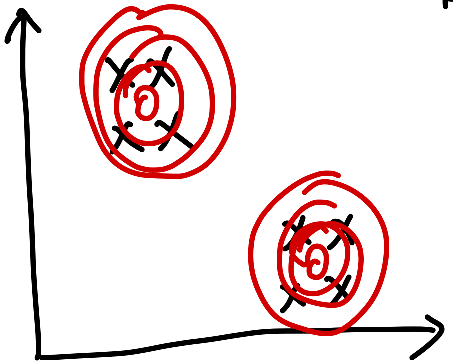


C_k = cluster $\#k$

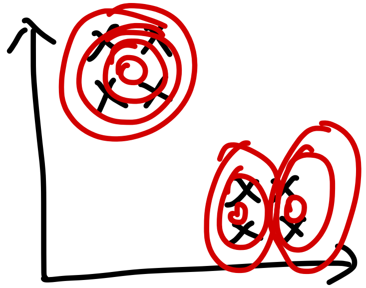
$$p(x) = \sum_{k=1}^{K} \underbrace{p(x \mid C_k)}_{\text{component densities}} \underbrace{Pr(C_k)}_{\text{mixture proportions}}$$



$K=1$



$K=2$



$K=3$

K = # of components (clusters) (groups)

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

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

↳ cluster/group/component membership

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

Iterative algorithm

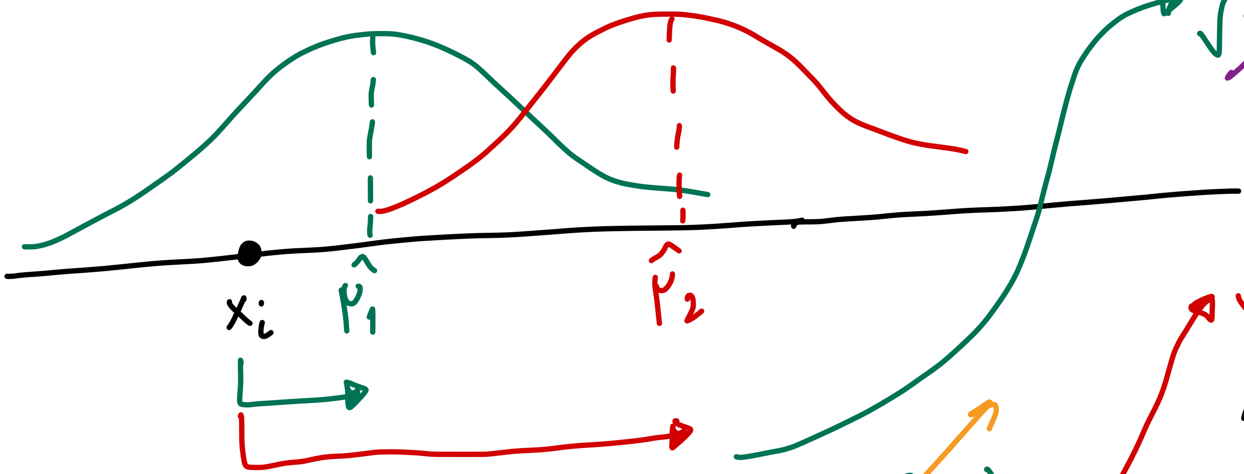
STEP 1: Estimate the cluster memberships ($\hat{y}_{ik}$)

STEP 2: Estimate the parameters.

$$\hat{Pr}(C_k) = \frac{\sum_{i=1}^{N} \hat{y}_{ik}}{N} \qquad \hat{\mu}_k = \frac{\sum_{i=1}^{N} \hat{y}_{ik} . x_i}{\sum_{i=1}^{N} \hat{y}_{ik}} \qquad \hat{\Sigma}_k = \frac{\sum_{i=1}^{N} \hat{y}_{ik}(x_i-\hat{\mu}_k)(x_i-\hat{\mu}_k)^T}{\sum_{i=1}^{N} \hat{y}_{ik}}$$

center of mass

K-MEANS CLUSTERING



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

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

Assuming $\underline{Pr(C_1) = Pr(C_2)}$

& $\underline{\sigma_1^2 = \sigma_2^2}$

$$Pr(C_1 \mid x) = \frac{p(x \mid C_1)\, \cancel{Pr(C_1)}}{p(x)}$$

$$Pr(C_2 \mid x) = \frac{p(x \mid C_2)\, \cancel{Pr(C_2)}}{p(x)}$$

Compare $\|x_i-\hat{\mu}_1\|_2$ and $\|x_i-\hat{\mu}_2\|_2$

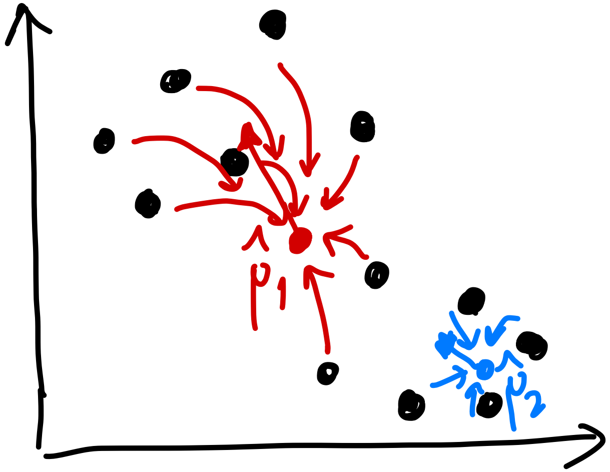
$$\left.\begin{array}{l} y_{i1}=0 \\ y_{i2}=1 \end{array}\right\} \text{if } \|x_i-\hat{\mu}_1\|_2 > \|x_i-\hat{\mu}_2\|_2$$

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

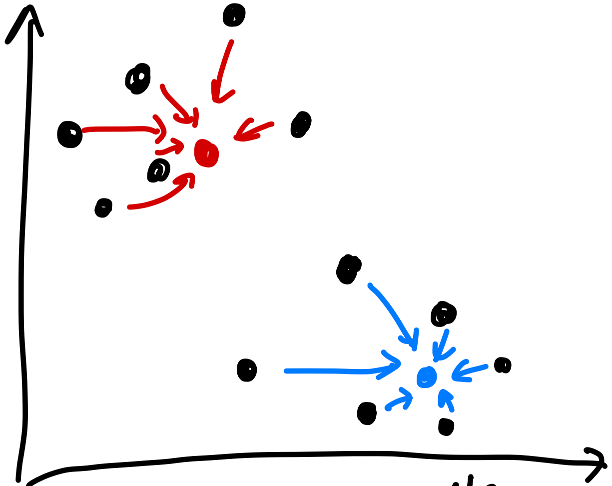
$$b_{ik} = \begin{cases} 1 & \text{if } \|x_i - \hat{\mu}_k\|_2 = \min_{c=1}^{K} \|x_i - \hat{\mu}_c\|_2 \\ 0 & \text{otherwise} \end{cases}$$

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

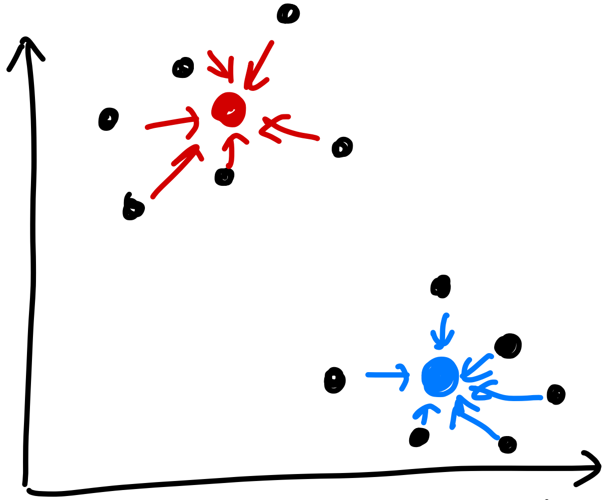
- Initialize $\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_K$ randomly
- Repeat
  - for all x_i:
    $$b_{ik} = \begin{cases} 1 & \text{if } \|x_i - \hat{\mu}_k\|_2 = \min_{c=1}^{K} \|x_i - \hat{\mu}_c\|_2 \\ 0 & \text{otherwise} \end{cases}$$
  - for all $\hat{\mu}_k$:
    $$\hat{\mu}_k = \frac{\sum_{i=1}^{N} b_{ik} \cdot x_i}{\sum_{i=1}^{N} b_{ik}}$$
- Until convergence [[all b_{ik}'s stay the same]] or [[all $\hat{\mu}_k$'s stay the same]]



Iteration #1



Iteration #2



Iteration #3

b_{ik}'s stayed the same