Next week Exam 2 : 12/10 → Friday 12/13

(CFE) comprehensive final Exam

@ 11 am

Top 7

grading policy

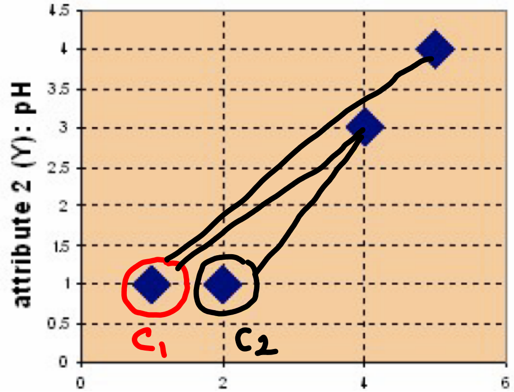
HW 1
2
3
4
5
6
7
8

HW / Quiz	Exam 1	Exam 2	C.F.E	FTP
25%	25%	25%	25%	25%

Top 3 (75%) + 25% = 100%

Object	weight Index	pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

$A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad B = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad C = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \quad D = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$



1_ Let $K=2$

2_ Initialize the centroid $c_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $c_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

3_ Distance

$$D^0 = \begin{bmatrix} 0 & 1 & \sqrt{(4-1)^2+(3-1)^2} & \sqrt{(5-1)^2+(4-1)^2} \\ 1 & 0 & \sqrt{(4-2)^2+(3-1)^2} & \sqrt{(5-2)^2+(4-1)^2} \end{bmatrix}$$

$$D^0 = \begin{bmatrix} 0 & 1 & 3.6 & 5 \\ 1 & 0 & 2.8 & 4.2 \end{bmatrix} \begin{matrix} \text{class 1} \\ \text{class 2} \end{matrix}$$

4- $\boxed{\text{object clustering assignment}}$: we assign each object based on minimum distance

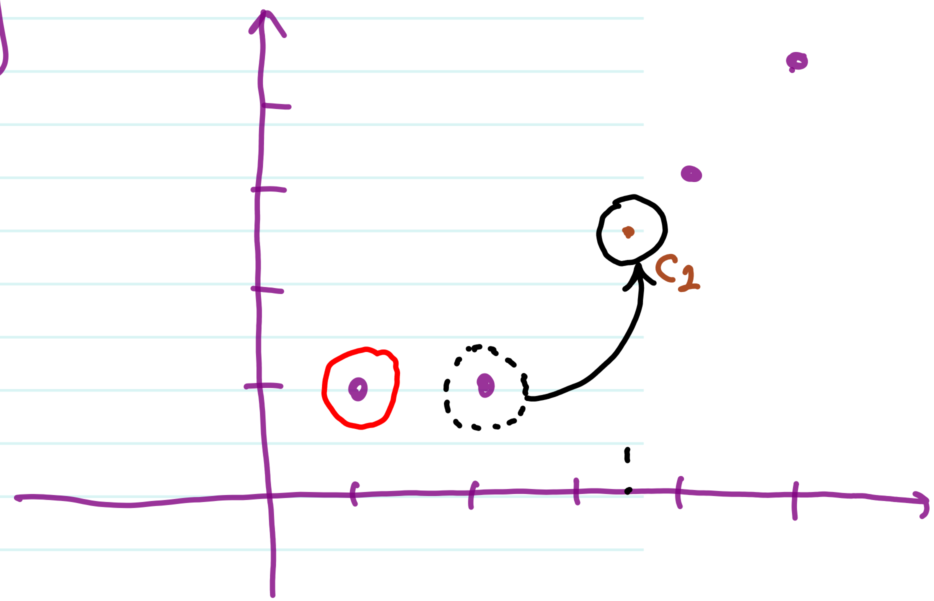
$$G^0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

$A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $B = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ $C = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ $D = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$

5- update centroid:

c_1 will not change $= \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$c_2 = \begin{bmatrix} \frac{2+4+5}{3} \\ \frac{1+3+4}{3} \end{bmatrix} = \begin{bmatrix} 11/3 \\ 8/3 \end{bmatrix}$$



6- Repeat process

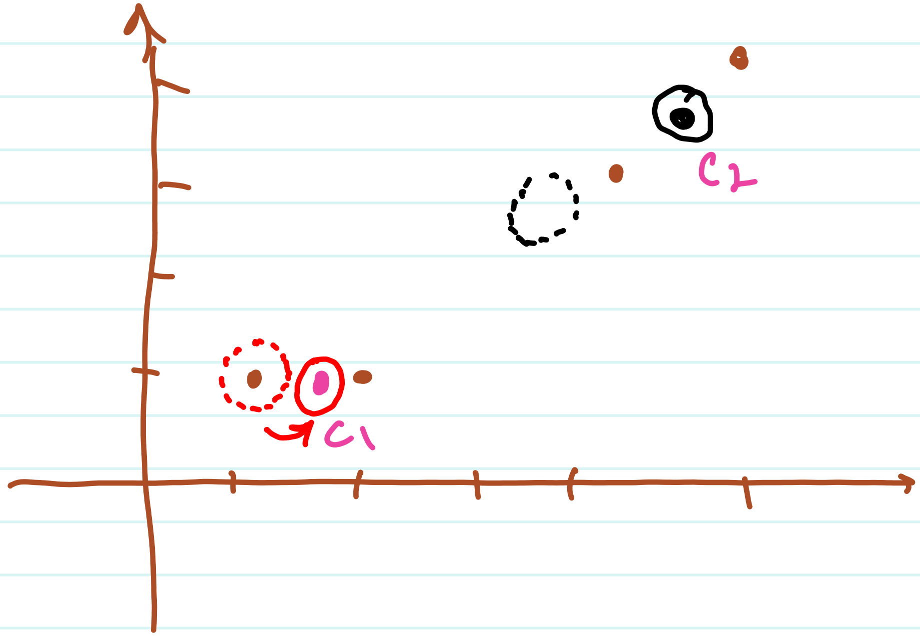
$$D^1 = \begin{bmatrix} 0 & 1 & 3.6 & 5 \\ \sqrt{\left(\frac{11}{3}-1\right)^2+\left(\frac{8}{3}-1\right)^2} & \sqrt{\left(\frac{11}{3}-2\right)^2+\left(\frac{8}{3}-1\right)^2} & \sqrt{\left(\frac{11}{3}-4\right)^2+\left(\frac{8}{3}-3\right)^2} & \sqrt{\left(\frac{11}{3}-5\right)^2+\left(\frac{8}{3}-4\right)^2} \end{bmatrix}$$

$$D^1 = \begin{bmatrix} 0 & 1 & 3.6 & 5 \\ 3.14 & 2.35 & 0.47 & 1.88 \end{bmatrix}$$

$$G^1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

calculate centroid $\quad$ data $= \begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 1 & 3 & 4 \end{bmatrix}$

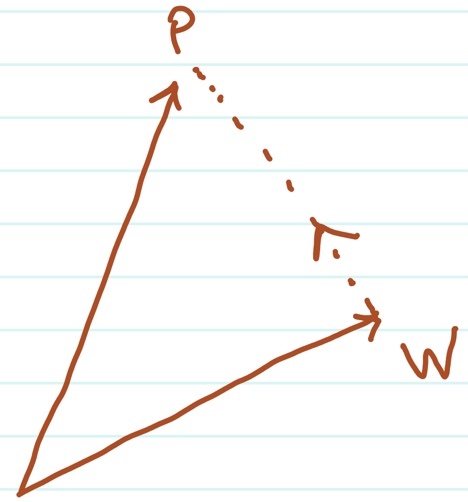
$$C_1 = \begin{bmatrix} \frac{1+2}{2} \\ \frac{1+1}{2} \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1 \end{bmatrix} \qquad C_2 = \begin{bmatrix} \frac{4+5}{2} \\ \frac{3+4}{2} \end{bmatrix} = \begin{bmatrix} 4.5 \\ 3.5 \end{bmatrix}$$



$$D^2 = \begin{bmatrix} \sqrt{(1.5-1)^2+(1-1)^2} & \sqrt{(1.5-2)^2+(1-1)^2} & \sqrt{(1.5-4)^2+(1-3)^2} & \sqrt{(1.5-5)^2+(1-4)^2} \\ \sqrt{(4.5-1)^2+(3.5-1)^2} & \sqrt{(2-4.5)^2+(3.5-1)^2} & \sqrt{(4-4.5)^2+(3-3.5)^2} & \sqrt{(5-4.5)^2+(3.5-4)^2} \end{bmatrix}$$

$$D^2 = \begin{bmatrix} 0.5 & 0.5 & 3.2 & 4.6 \\ 4.3 & 2.91 & 0.7 & 0.7 \end{bmatrix} \rightarrow G^2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$G^2 = G^1 \rightarrow$ Converagne $\rightarrow$ $\boxed{\text{stop}}$


P
r
w

Cluster Label

Point	Cluster Label
P1	1 ✓
P2	1 ✓
P3	2 ✓
P4	2 ✓

class I

class II

Dissimilarity Matrix

MORE VIDEOS

Point	P1	P2	P3	P4
P1	0	0.10	0.65	0.55
P2	0.10	0	0.70	0.60
P3	0.65	0.70	0	0.30
P4	0.55	0.60	0.30	0

$$SC = 1 - \frac{a}{b}$$

b

$P_1 \to P_3$

$P_1 \to P_4$

$\boxed{\text{point } 1}$

$$a = \frac{0.1}{1}$$

$$b = \frac{0.65 + 0.55}{2} = 0.6$$

$$SC = 1 - \frac{a}{b} = 1 - \frac{0.1}{0.6} = \underline{0.83}$$

point 2

a

$P_2 \longrightarrow P_1$

b

$P_2 \rightarrow P_3$

$P_2 \rightarrow P_4$

$a = 0.1$

$b = \frac{0.7 + 0.6}{2} = 0.65$

$Sc = 1 - \frac{0.1}{0.65} = \underline{0.846}$

point 3

a

$P_3 \longrightarrow P_4$

b

$P_3 \rightarrow P_1$

$P_3 \rightarrow P_2$

$Sc = 1 - \frac{0.3}{\frac{0.65 + 0.7}{2}} = \underline{0.556}$

Poin 4

a

$P_4 \longrightarrow P_3$

b

$P_4 \rightarrow P_1$

$P_4 \rightarrow P_2$

$Sc = 1 - \frac{0.3}{\frac{0.55 + 0.6}{2}} = \underline{0.478}$

	SC
P_1	0.83
P_2	0.846
P_3	0.556
P_4	0.478

cluster 1 : $\frac{0.83 + 0.846}{2} = 0.84$

cluster 2 : $\frac{0.556 + 0.478}{2} = 0.517$

overal SC $= \frac{0.84 + 0.517}{2} = 0.68$

$$Sc = 1 - \frac{a}{b} = 1 \rightarrow a = 0$$

$$1 - \frac{a}{b} = 0 \rightarrow a = b$$

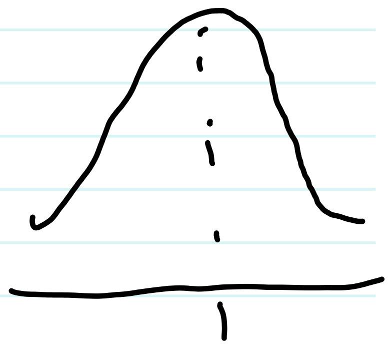
$$1 - \frac{a}{b} = -1 \rightarrow \frac{a}{b} = 2 \rightarrow a = 2b$$

density based Algorithm:

Expected Maximization Algorithm: $\sim$ MLE

Tell: $[1, 2, x]$ Are drawn from $N(1, 1)$

$x \downarrow 1$



Tell: $[0, 1, 2]$ Are drawn from $N(\mu, 1)$

$$\hat{\mu} = \frac{0+1+2}{3} = 1$$

Now Tell

$[1, 2, x]$ Are drawn from $\mathcal{N}(\mu, 1)$

EM- Algorithm

guess $\mu_0 = 0 \longrightarrow x_0 = 0$

$[1, 2, 0] \longrightarrow \mu_1 = \dfrac{1+2+0}{3} = 1 \longrightarrow x_1 = 1$

$[1, 2, 1] \longrightarrow \mu_2 = \dfrac{1+2+1}{3} = 4/3 \longrightarrow x_2 = 4/3$

.... continue.

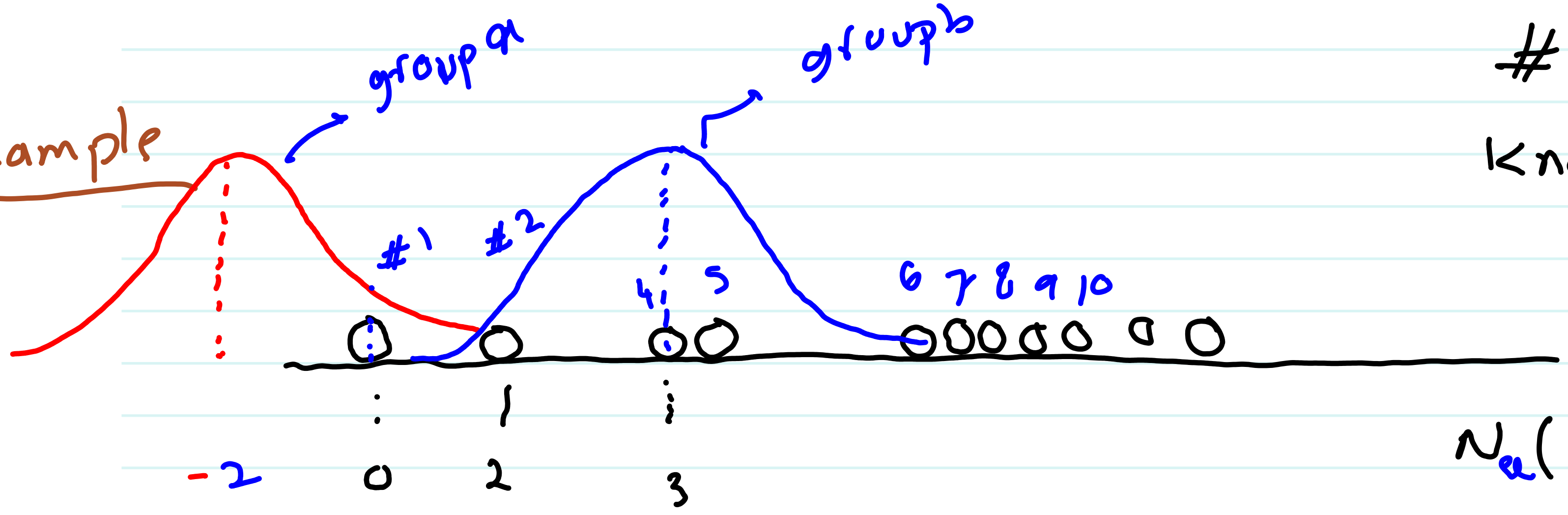
$$\boxed{x^* = \mu^* = 1.5}$$

clusting Gaussian mixture Model (GMM)

$\#$ 10 observation

Know $K=2 \begin{cases} \text{class } a \\ \text{class } b \end{cases}$

Example



$N_a(\mu_1, 1)$

$N_b(\mu_2, 1)$

$\theta = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}$

Use EM Algorithm to find μ_1, μ_2

1- make initial guess: $\mu_1 = -2$

$\mu_2 = 3$

$b_i = 1 - a_i$

$$\text{prob}(x_i \mid a) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \mu_a)^2}{2}}$$

$$a_i = P(a \mid x_i;\theta) = \frac{P(x_i \mid a)\cdot P(a)}{P(x_i \mid a)\cdot P(a) + P(x_i \mid b)\cdot P(b)}$$

$$P(0 \mid a,\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(2)^2}{2}} = 5\%$$

$P(a) = P(b) = 0.5$

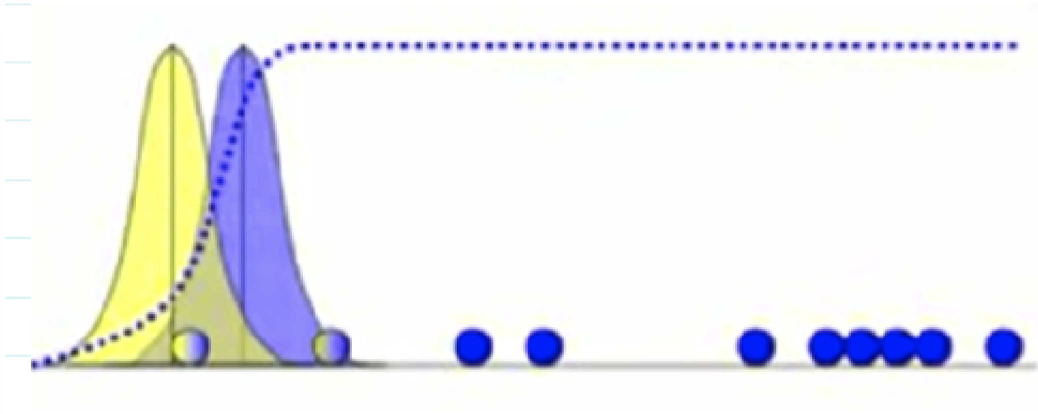
$$P(0 \mid b,\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(-3)^2}{2}} = 0.4\%$$

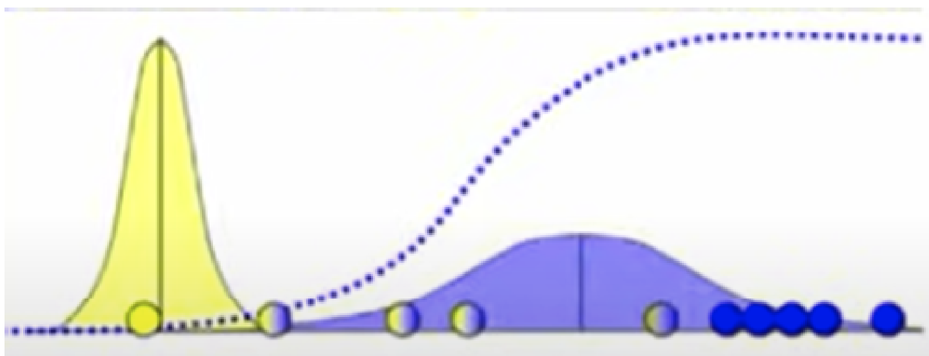
$$P(a \mid 0,\theta) = \frac{P(0 \mid a)\cdot \cancel{P(a)}}{P(0 \mid a)\cdot \cancel{P(a)} + P(0 \mid b)\cdot \cancel{P(b)}} = \frac{5\%}{5\% + 0.4\%} = 92\%$$

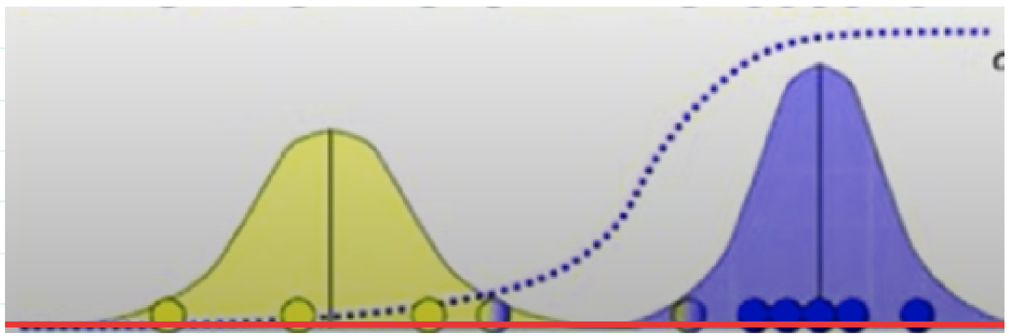
$$P(b \mid 0,\theta) = 1 - 92\% = 8\%$$

$p(1 \mid a, \theta) =$

$p(a \mid 1, \theta)$ $p(b \mid 1, \theta)$







update mean

$$\mu_a = \frac{a_1 x_1 + a_2 x_2 + \cdots + a_n x_n}{a_1 + \cdots + a_n}$$

variance

$$\sigma_a^2 = \frac{a_1 (x_1 - \mu_a)^2 + \cdots + a_n (x_n - \mu_a)^2}{a_1 + \cdots + a_n}$$

$$\mu_b = \frac{b_1 x_1 + \cdots + b_n x_n}{b_1 + \cdots + b_n} \qquad \sigma_b^2 = \frac{b_1 (x_1 - \mu_b)^2 + \cdots + b_n (x_n - \mu_b)^2}{\sum_{i=}^{n} b_i}$$

$$P(a) = \frac{a_1 + a_2 + \cdots + a_n}{n}, \quad P(b) = 1 - P(a)$$

better estimate of prior probality.

Suppose this is our dataset of any supermarket, where user id and items are as shown below. Using apriori algorithm find the strongest association rule between items.

User ID	Items
001	1, 3, 4
002	2, 3, 5
003	1, 2, 3, 5
004	2, 5

Step 1: set min support and confindence

min sup = 50%   min conf = 70%

Step 2: calculate support for each item 1, 2, 3, 4, 5

1	2/4 = 50%
2	75%
3	75%
4	25%
5	75%

remove item 4   25% < 50%

$\Downarrow$

keep {1, 2, 3, 5}

Step3

item	support
1, 2	~~25%~~
1, 3	50%
1, 5	~~25%~~
2, 3	50%
2, 5	75%
3, 5	50%

min sup is 50%

remove

{1, 2} , {1, 5}

keep

{1, 3}, {2, 3}, {2, 5} {3, 5}

Step4

item	support
~~{1, 3, 2}~~	~~25%~~
~~{1, 3, 5}~~	~~25%~~
{2, 3, 5}	50%
~~{1, 2, 5}~~	~~25%~~

remove the red

keep {2, 3, 5}

Step 5 Form association rule {2, 3, 5}

$$C(A \to B) = \frac{sup(A,B)}{sup(A)}$$

rules	support count	confidence
$\boxed{\{2,3\} \to \{5\}}$	2	$\boxed{\frac{2/4}{2/4} = 100\%}$
~~{2,5} → {3}~~	2	~~$\frac{2/4}{3/4} = 66\%$~~
$\boxed{\{3,5\} \to \{2\}}$	2	$\boxed{2/4 / 2/4 = 100\%}$
~~{2} → {3,5}~~	2	~~66%~~
~~{3} → {2,5}~~	2	~~66%~~
~~{5} → {3,2}~~	2	~~66%~~

considering 70% min conf.

min sup 50% min conf 70%

$\boxed{\{2,3\} \to \{5\}}$

$\{3, 5\} \to \{2\}$