21-S1-Q4

Q: (a) K mean

(b) PCA

(c) Bayes

Solution (a) ① Initialize : $C_1 = 3$ $C_2 = 13$

② Iterate 1

Assign every data point to its closest cluster center, according to Euclidean distance

		5	7	10	12
3	C_1	2	4	7	9
13	C_2	8	6	3	1
	Assign	C_1	C_1	C_2	C_2

Update the cluster center.

C_1 : $\frac{5+7}{2} = 6$

C_2 : $\frac{10+12}{2} = 11$

③ Iteration 2

		5	7	10	12
6	C_1	1	1	4	6
11	C_2	6	4	1	1
	Assign	C_1	C_1	C_2	C_2

④ Stopping : no points' assignments change

(b) (1) ① $\mu = (2, 2)$

② center the data

$(-1, -1)$ $(0, 0)$ $(1, 1)$

③ Covariance matricx

$\sigma_{11}^2 = \frac{1}{3} \times [(-1)^2 + 0^2 + 1^2] = \frac{2}{3}$

$\sigma_{22}^2 = \frac{2}{3}$

$\sigma_{12}^2 = \frac{2}{3} = \sigma_{21}^2$

$$\Sigma = \begin{bmatrix} \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix}$$

④ Eigen-decomposition

let $|\lambda I - \Sigma| = \begin{vmatrix} \lambda - \frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \lambda - \frac{2}{3} \end{vmatrix}$

$= (\lambda - \frac{2}{3})^2 - (\frac{2}{3})^2$

$= 0$

$\lambda^2 - \frac{4}{3}\lambda + \frac{4}{9} = \frac{4}{9}$

$\lambda(\lambda - \frac{4}{3}) = 0$

$\lambda_1 = \frac{4}{3}$ $\lambda_2 = 0$

when $\lambda_1 = \frac{4}{3}$

$$(\frac{4}{3} I - \Sigma) x = \begin{bmatrix} \frac{4}{3} - \frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} - \frac{2}{3} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= 0$$

$$\begin{bmatrix} \frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} \end{bmatrix} -1 \rightarrow \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} \\ 0 & 0 \end{bmatrix} \times \frac{3}{2} \rightarrow \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$$

let $x_2 = k$

$x_1 - x_2 = 0 \implies x_1 = x_2 = k$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} k \\ k \end{bmatrix} = k \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

So eigenvector $v_1 = k \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

unit ~ $v_1 = \begin{bmatrix} \frac{1}{\sqrt{1+1}} \\ \frac{1}{\sqrt{1+1}} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$

(i) First PC : $v_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$

(ii) variance after 1-D projection

$\lambda_1 = \frac{4}{3} = 1.3333$

(c) (i)

$$p(E) = \frac{4}{6 \times 6}$$

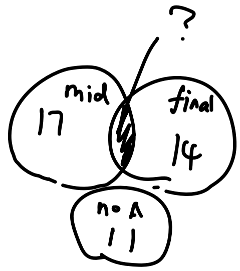
$$= \frac{1}{9}$$

14
23
32
41

(ii) $35 - 11 = 24$

$17 + 14 - 24 = 7$

$$p(\text{both}) = \frac{7}{35} = 0.2$$



(iii) $p(\text{COVID}) = 0.001$

$p(\text{positive} \mid \neg\text{COVID}) = 0.005$

$p(\text{COVID} \mid \text{positive}) = ?$

$$p(\text{COVID} \mid \text{positive}) = \frac{p(\text{positive} \mid \text{COVID}) \times p(\text{COVID})}{p(\text{positive} \mid \text{COVID}) \times p(\text{COVID}) + p(\text{positive} \mid \neg\text{COVID})\, p(\neg\text{COVID})}$$

$$= \frac{1 \times 0.001}{1 \times 0.001 + 0.005 \times 0.999}$$

$$= 0.167$$