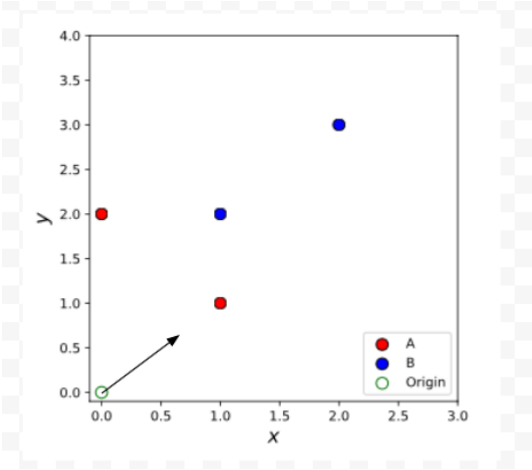
Q1:

a)



$$m = \frac{1}{4}\left(\begin{bmatrix}0\\2\end{bmatrix}+\begin{bmatrix}1\\1\end{bmatrix}+\begin{bmatrix}1\\2\end{bmatrix}+\begin{bmatrix}2\\3\end{bmatrix}\right) =$$

$$= \begin{bmatrix}1\\2\end{bmatrix}$$

$$Cov(X) = \frac{1}{n-1} XX^T = \frac{1}{3}\begin{bmatrix}-1 & 0 & 0 & 1\\ 0 & -1 & 0 & 1\end{bmatrix}\begin{bmatrix}-1 & 0\\ 0 & -1\\ 0 & 0\\ 1 & 1\end{bmatrix} =$$

$$= \begin{bmatrix}\frac{2}{3} & \frac{1}{3}\\ \frac{1}{3} & \frac{2}{3}\end{bmatrix}$$

$$\det(Cov(x) - \lambda I) = 0$$

$$\det\left(\begin{bmatrix}\frac{2}{3}-\lambda & \frac{1}{3}\\ \frac{1}{3} & \frac{2}{3}-\lambda\end{bmatrix}\right) = 0$$

$\lambda_1 = 1$ $\quad w_1 =$ [illegible] $\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$

$\lambda_2 = \frac{1}{3}$ $\quad w_2 =$ [illegible] $\begin{bmatrix}\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}\end{bmatrix}$

[illegible]

Assume that [illegible]

c) $S_W = S_A + S_B = \sum_{i=1}^{2} \sum_{j=1}^{2} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T =$

$= \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{bmatrix} + \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{bmatrix} + \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} + \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} =$

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

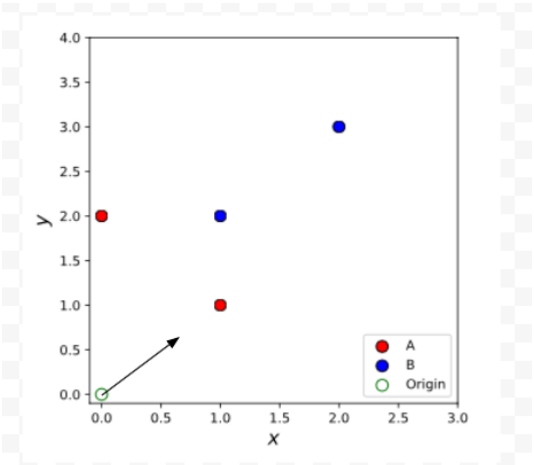
$S_B = (\bar{x}_A - \bar{x}_B)(\bar{x}_A - x_B)^T = 2$

d) $w = c \cdot S_W^{-1}(m_A - m_B)$

$m_A - m_B = \begin{bmatrix} \frac{1}{2} \\ \frac{3}{2} \end{bmatrix} - \begin{bmatrix} \frac{3}{2} \\ \frac{5}{2} \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

$w = c \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \end{bmatrix} = c \cdot \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

$w = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$


4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
y
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x
A
B
Origin

$\sqrt{2}$

a) The

1st iteration:

$r_{11} = 1, r_{21} = 1, r_{31} = 1$

$r_{42} = 1, r_{52} = 1, r_{62} = 1, r_{72} = 1, r_{82} = 1$

cluster centers:

$$c_1 = \frac{x_1 + x_2 + x_3}{3} = 3\frac{2}{3}$$

$$c_2 = \frac{x_4 + x_5 + x_6 + x_7 + x_8}{5} = 10$$

2nd iteration:

$r_{11} = 1, r_{21} = 1, r_{31} = 1, r_{41} = 1$

$r_{52} = 1, r_{62} = 1, r_{72} = 1, r_{82} = 1$

cluster centers:

$$c_1 = \frac{x_1 + x_2 + x_3 + x_4}{4} = 4.25$$

$$c_2 = \frac{x_5 + x_6 + x_7 + x_8}{4} = 11$$

3rd iteration:

$r_{11} = 1, r_{21} = 1, r_{31} = 1, r_{41} = 1$

$r_{51} = 1, r_{61} = 1, r_{72} = 1, r_{82} = 1$

As there is no change in cluster assignment.
The cluster centers are final: $c_1 = 4.25$

$c_2 = 11$

b) It takes 3 iterations

$$J = \sum_{n=1}^{8} \sum_{k=1}^{2} r_{nk} \|x_n - \mu_k\|^2 =$$

$$= (2-4.25)^2 + (4-4.25)^2 + (5-4.25)^2 +$$

$$+ (8-11)^2 + (10-11)^2 + (12-11)^2 + (14-11)^2 =$$

$$= 28.75$$

c) 1st iteration:

$r_{11}=1, r_{21}=1, r_{31}=1, r_{41}=1, r_{51}=1$

$r_{62}=1, r_{72}=1, r_{82}=1$

$$\mu_1 = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = 5$$

$$\mu_2 = \frac{x_6 + x_7 + x_8}{3} = 12$$

2nd iteration:

$r_{11}=1, r_{21}=1, r_{31}=1, r_{41}=1, r_{51}=1$

$r_{62}=1, r_{72}=1, r_{82}=1$

As there is no change in cluster assignment

The final cluster centers: $\mu_1 = 5$

$\mu_2 = 12$

d) It takes 2 iterations

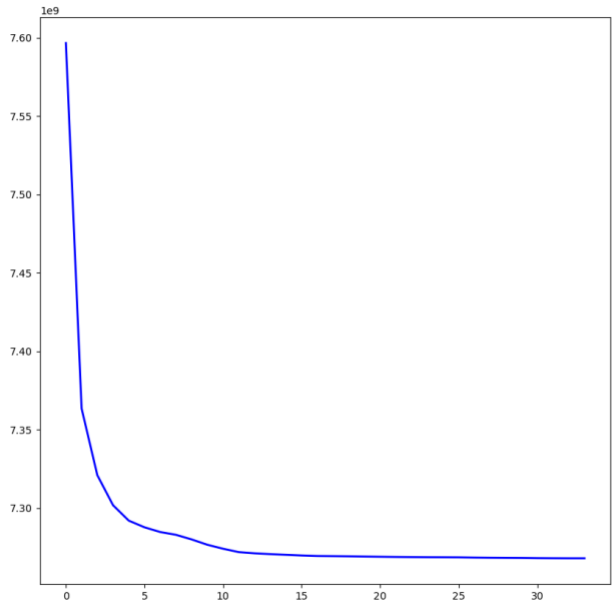
$$J = \sum_{n=1}^{8} \sum_{k=1}^{2} r_{nk} \| x_n - \mu_k \|^2 =$$

$$= (2-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (8-5)^2 +$$
$$+ (10-12)^2 + (12-12)^2 + (14-12)^2 = 28$$

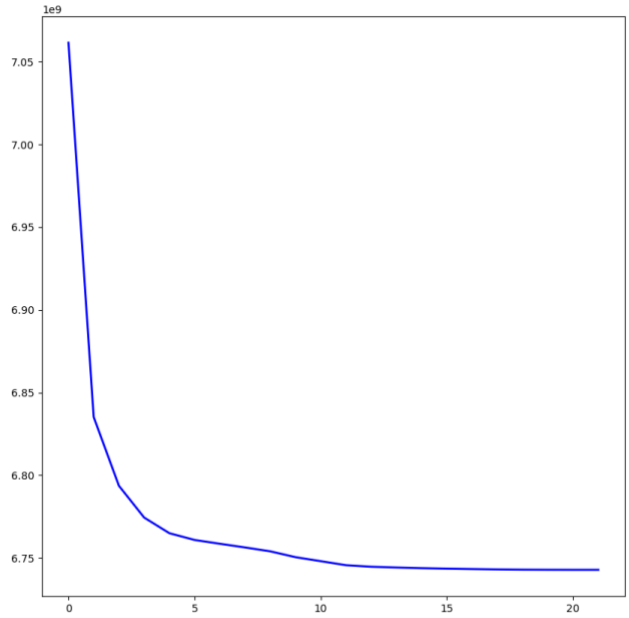
e) It takes one iteration less for c than for a
And $J_c < J_a$. Hence, c is better

Q3:

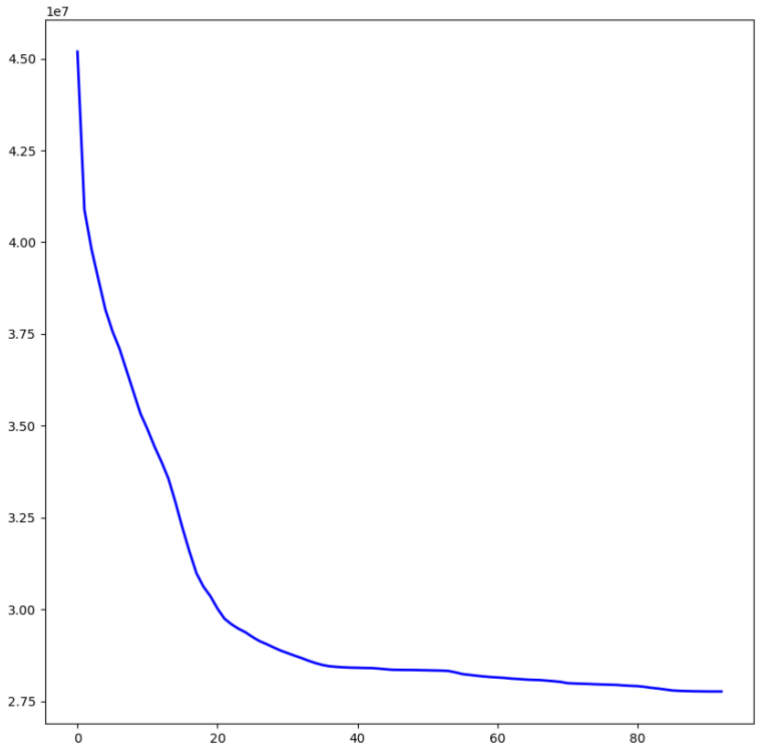
a) entropy = 0.367, 34 iterations.



b) 128 dimensions. PCA helps clustering. Reducing the number of dimensions reduces the number of iterations and run-time (run-time: 10.26 part (a), 5.62 part(b); number of iterations: 34 part (a), 22 part(b)). Entropy 0.373



c) The result is worse than for part (b). The number of iterations is 93, entropy is 0.674, run-time is 22.06. PCA from part (b) makes the number of dimensions such that it captures >95% of the variance. While PCA from part (c) makes only 1 dimension which changes variance significantly.


1e7
4.50
4.25
4.00
3.75
3.50
3.25
3.00
2.75
0
20
40
60
80