Programming 3 | CS 545 Summer 2023 | Leo Lu

1.Let's assume that there are N doors, and that the host can open G number of doors.

Probability of selecting the car at the start: 1/N

Probability of selecting a goat at the start : (N-1)/N

After G doors are revealed, N-G-1 doors will remain unselected.

Therefore, if switching, the probability of selecting the car becomes (N-1)/ N(N-G-1)

P(Select Car before reveal) = $\frac{1}{4}$
P(Select Goat 1 before reveal) = $\frac{1}{4}$
P(Select Goat 2 before reveal) = $\frac{1}{4}$
P(Select Goat 3 before reveal) = $\frac{1}{4}$

Switching, probability of selecting car: (4-1)/ 4(4-1-1) = $\frac{3}{8}$
Not switching, probability of selecting car = P(Select car before reveal) = $\frac{1}{4}$ = 2/8
Therefore, it is advisable to switch doors since $\frac{3}{8}$ > 2/8 (keeping same door) .

2a.
C: Cough
F: Fever
Ch: Chills
Co: Congestion

Classes
A : Flu
B : Cold
P(A) = 4/6
P(B) = 2/6

P(C|A)= 2/4, P(!F|A)=$\frac{1}{4}$, P(!Ch|A)=$\frac{1}{4}$, P(Co|A) = $\frac{1}{4}$
P(C|B)= $\frac{1}{2}$, P(!F|B)=$\frac{1}{2}$, P(!Ch|B)=$\frac{1}{2}$, P(Co|B) = 2/2
Using the Naive Bayes Classifier:

P(A|C, !F, !Ch, Co) ~= P(C|A) x P(!F|A) x P(!Ch|A) x P(Co|A) x P(A)
= 2/4 x $\frac{1}{4}$ x $\frac{1}{4}$ x $\frac{1}{4}$ x 4/6 = $\frac{1}{2}$ x $\frac{1}{4}$ x $\frac{1}{4}$ x $\frac{1}{4}$ x $\frac{2}{3}$ = 2/384 = 1/192

P(B|C, !F, !Ch, Co) ~= P(C|B) x P(!F|B) x P(!Ch|B) x P(Co|B) x P(B)
= $\frac{1}{2}$ x $\frac{1}{2}$ x $\frac{1}{2}$ x 2/2 x 2/6 = $\frac{1}{8}$ x $\frac{1}{3}$ = 1/24

Based on the above 1/24 > 1/192, it is more probable that patient P7 has the Cold than having the Flu.

2b.
P(Congestion=True | Cold) = 2/2 ~= (2+1)/(2+2) = $\frac{3}{4}$

P(Congestion False | Cold) = 0/2 ~= 1/ (2+2) = ¼

3.

$w \cdot x = \beta_0 + \beta_1 x$

$P(Y = 1|x) = \frac{1}{1+e^{-w \cdot x}}$

$P(Y = 0|x) = \frac{e^{-w \cdot x}}{1+e^{-w \cdot x}}$

P(no bacteria | dosage = 1.0) ~ P(Y=1 | x=1.0) = 1/ (1+ e^1.621) = 0.1650670041379 ~= 0.165

P(no bacteria | dosage = 2.5) ~ P(Y=1 | x=2.5) = 1/(1+e^-3.818) = 0.9785006763174 ~= 0.979

P(no bacteria | dosage = 2.5) ~ P(Y=0 | x=2.0) = e^-2.005/(1+e^-2.005) = 0.1186789528164 ~= 0.119

ii)

P(Y=1 | x=?) = 0.8

$\frac{1}{1+e^{-(-5.247+3.626x)}} = 0.8$

$1 + e^{5.247-3.626x} = 1.25$

$e^{5.247-3.626x} = 0.25$

$5.247 - 3.626x = ln(0.25)$

$3.626x = 5.247 - ln(0.25)$

$3.626x = 5.247 - (-1.3862943611199)$

$3.626x = 6.6332943611199$

$x = 1.8293696528185 \approx 1.829$

The dosage is 1.829mg for the 80th percentile cutoff producing no bacteria.

4.

System:

$$x + y = 1$$
$$x + 2y = 2$$
$$x + 3y = 2$$
$$x + 4y = 3$$

Using OLS

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{bmatrix}$$

$$(A^T A)^{-1} = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \right)^{-1} = \begin{pmatrix} 4 & 10 \\ 10 & 30 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{5} \end{pmatrix}$$

$$x^* = (A^T A)^{-1} A^T b = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 2 \\ 3 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{3}{10} & -\frac{1}{10} & \frac{1}{10} & \frac{3}{10} \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{3}{5} \end{pmatrix}$$

5.i)

$$A = U\Sigma V^T$$
$$U^{-1}A(V^T)^{-1} = U^{-1}U\Sigma V^T(V^T)^{-1}$$
$$U^{-1}A(V^T)^{-1} = I\Sigma I = \Sigma$$

$$U^{-1}A(V^T)^{-1} = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}^{-1} \begin{pmatrix} 0 & 1 \\ 1 & \frac{3}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & \frac{3}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{4}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{1}{2\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

Solve for A

$$A = U\Sigma V^T = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{4}{\sqrt{5}} & \frac{1}{2\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1 \\ 1 & \frac{3}{2} \end{pmatrix}$$

ii)

Rank 1 approximation of $A = \sigma_1 u_1 v_1^T = 2\begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} = 2\begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{4}{5} \end{pmatrix} = \begin{pmatrix} \frac{2}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{8}{5} \end{pmatrix}$

iii)

$A^{-1} = (U\Sigma V^T)^{-1} = U^{-1}\Sigma^{-1}(V^T)^{-1}$

SVD for A^{-1}

$$U^{-1}\Sigma^{-1}(V^T)^{-1} = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}^{-1}\begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}^{-1}\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2\sqrt{5}} & \frac{4}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{17}{10} & -\frac{3}{5} \\ \frac{3}{5} & -\frac{4}{5} \end{pmatrix}$$

6a)

Based on the assignment step:

$$S_i^{(t)} = \left\{ x_p : \left\| x_p - m_i^{(t)} \right\|^2 \le \left\| x_p - m_j^{(t)} \right\|^2 \ \forall j, 1 \le j \le k \right\}$$

$||x_1 - m_1|| = \sqrt{(1-1)^2 + (1-1)^2 + (1-1)^2} = 0$

$||x_1 - m_2|| \ = \sqrt{(1-1)^2 + (1-2)^2 + (1-0)^2} = \sqrt{2}$

Since $||x_1 - m_1|| \ < \ ||x_1 - m_2||$, x_1 is in the m_1 cluster

$||x_2 - m_1|| \ = \sqrt{(1-1)^2 + (1-1)^2 + (4-1)^2} = \sqrt{9} = 3$

$||x_2 - m_2|| \ = \sqrt{(1-1)^2 + (1-2)^2 + (4-0)^2} = \sqrt{17} \approx 4.123$

Since $||x_2 - m_1|| \ < \ ||x_2 - m_2||$, x_2 is in the m_1 cluster

$||x_3 - m_1|| \ = \sqrt{(5-1)^2 + (1-1)^2 + (2-1)^2} = \sqrt{17} \approx 4.123$

$||x_3 - m_2|| \ = \sqrt{(5-1)^2 + (1-2)^2 + (2-0)^2} = \sqrt{21} \approx 4.583$

Since $||x_3 - m_1|| \ < \ ||x_3 - m_2||$, x_3 is in the m_1 cluster

$||x_4 - m_1|| \ = \sqrt{(1-1)^2 + (2-1)^2 + (0-1)^2} = \sqrt{2}$

$||x_4 - m_2|| \ = \sqrt{(1-1)^2 + (2-2)^2 + (0-0)^2} = 0$

Since $||x_4 - m_2|| \ < \ ||x_4 - m_1||$, x_4 is in the m_2 cluster

$$m_i^{(t+1)} = \frac{1}{\left|S_i^{(t)}\right|} \sum_{x_j \in S_i^{(t)}} x_j$$

Update Step:

$m_1 = (\frac{1}{3})[(1+1+5), (1+1+1), (1+4+2)] = (\frac{1}{3})(7,3,7) = (\frac{7}{3}, 1, \frac{7}{3})$

$m_2 = \frac{1}{1}(1,2,0) = (1,2,0)$

b)

$$Entropy(c_i) = -\sum_{j=1}^{|c_i|} p_{i,j} \log_2 p_{i,j}$$

Cluster m_1 classes: x_1 −> A , x_2 −> A, x_3 −> C

$Entropy(m_1) = -(\frac{2}{3}log_2\frac{2}{3} + \frac{1}{3}log_2\frac{1}{3}) = 0.91829583405 = 0.918$

Cluster m_2 classes: x_4 −> C

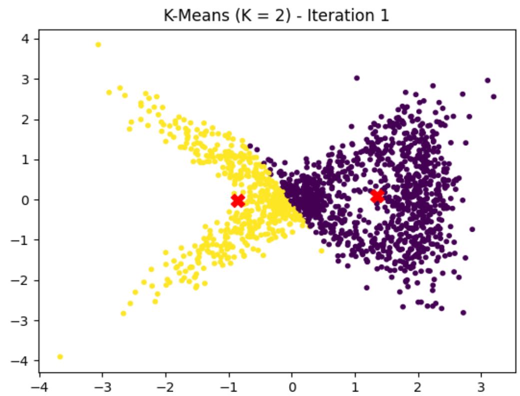
$Entropy(m_2) = -(0 + \frac{1}{1}log_2\frac{1}{1}) = 0$

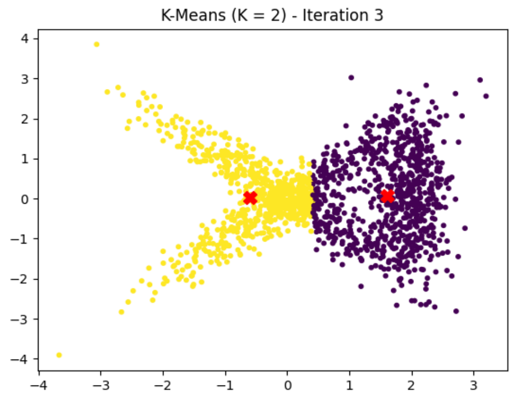
$$mean\ Entropy(C) = \sum_{i=1}^{K} \frac{n_i}{n} Entropy(c_i)$$

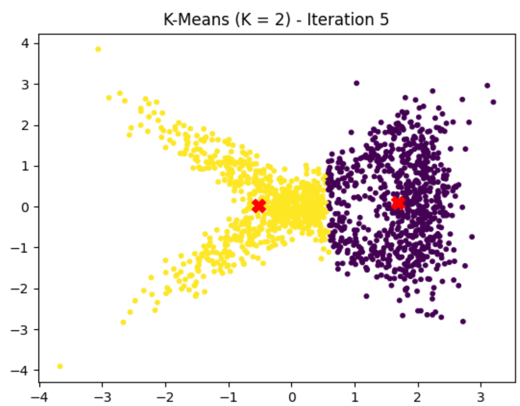
$$mean\ Entropy(m_1) \ = \ \frac{3}{4}(0.918) + \frac{1}{4}(0) = \frac{3}{4}(0.918) \ = \ 0.6885$$

Part 2

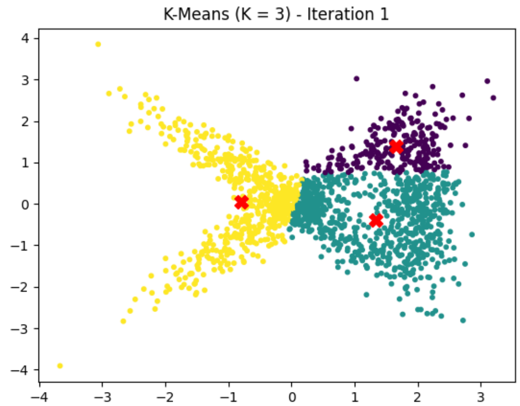
K=2

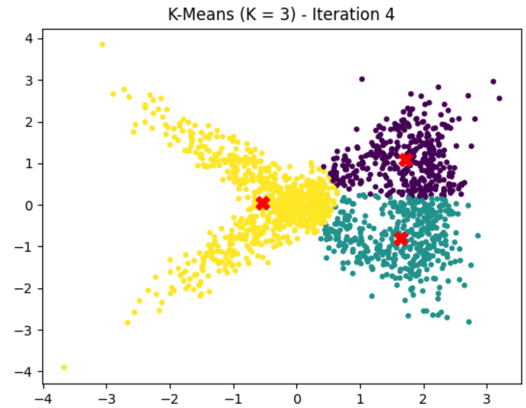


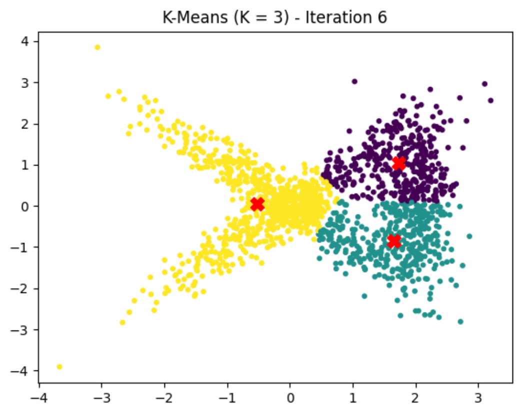


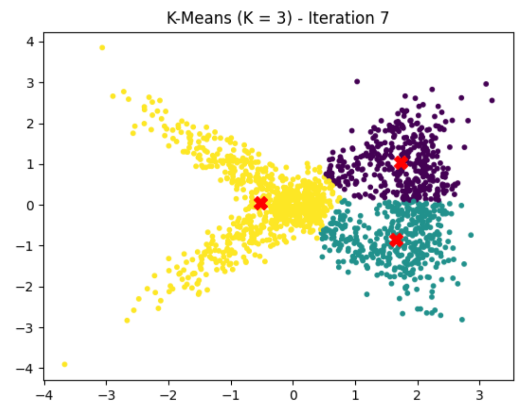


K=3

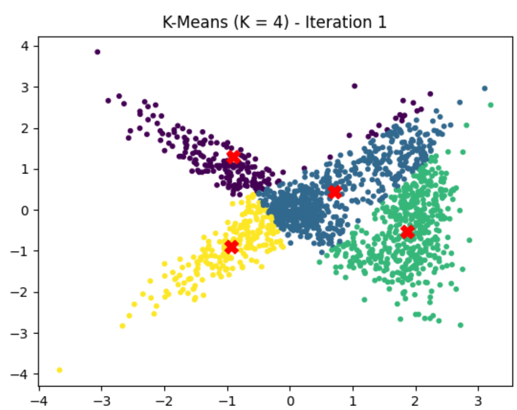

K-Means (K = 3) - Iteration 1

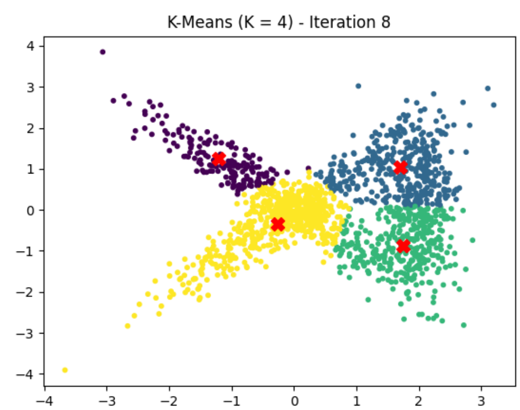

K-Means (K = 3) - Iteration 4

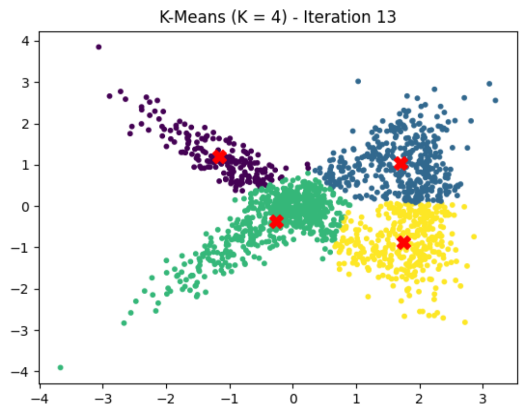

K-Means (K = 3) - Iteration 6

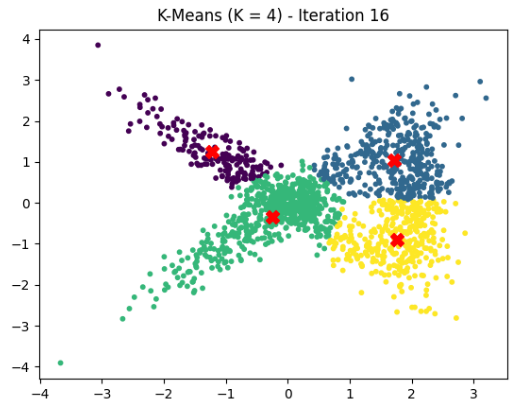

K-Means (K = 3) - Iteration 7

K=4

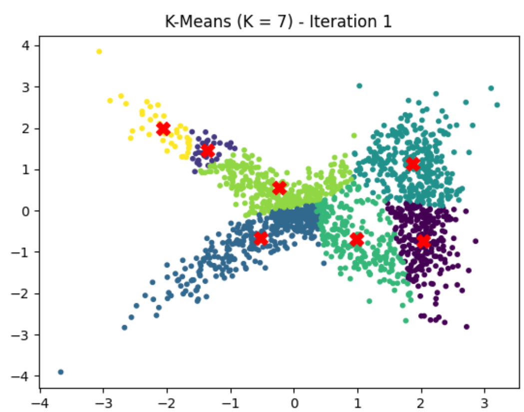

K-Means (K = 4) - Iteration 1

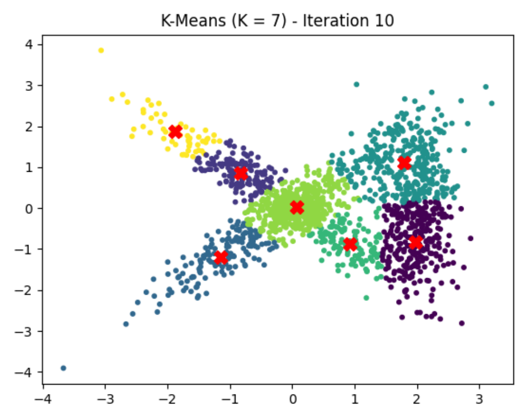

K-Means (K = 4) - Iteration 8

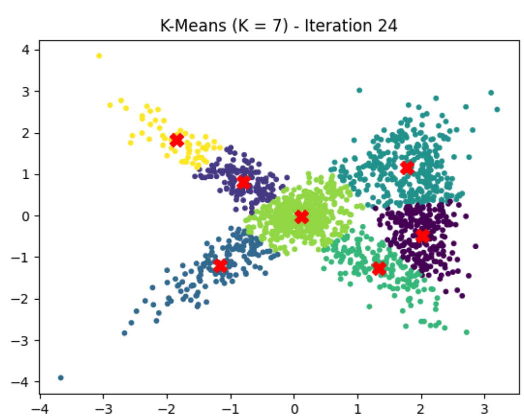


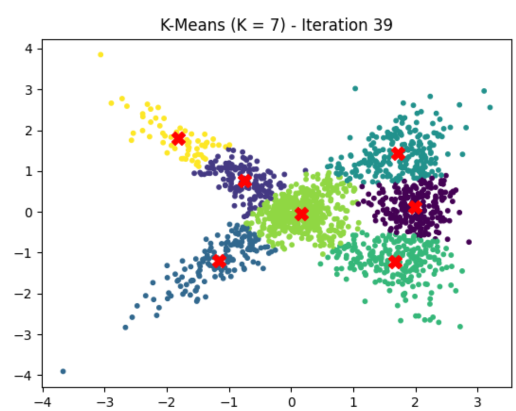


K=7









K	Sum of Squares Error

2	2228.6192
3	1539.2567
4	1103.5070
7	552.1652

Report

K-Means Clustering Performance Report

The K-Means algorithm was applied to a 2D dataset simulated from three Gaussians with considerable overlap. The goal was to cluster the data into K clusters for different values of K and analyze the resulting performance based on the sum of squares error and the number of iterations required for convergence. K-means was applied using 2, 3, 4, and 7 centroids with r=10.

Observations and Analysis

- As the value of K increases, the sum of squares error generally decreases, indicating better intra-cluster cohesion.
- Smaller values of K (2 and 3) lead to relatively higher errors, which might be due to the complexity of the data and overlap between clusters.
- A larger value of K (7) results in the lowest sum of squares error, suggesting a better fit to the data distribution.
- The number of iterations required for convergence varies with different values of K. Larger values of K tend to require more iterations for convergence.

Conclusion

The K-Means algorithm demonstrated varying levels of clustering performance for different values of K. The choice of K should be based on a trade-off between minimizing the sum of squares error and maintaining a meaningful number of clusters. The algorithm effectively captured the underlying structure of the data, demonstrating its utility in unsupervised learning tasks.