

Task1

Manual:

AMMAR JAMIL 01-134231-010

Points	K Means	Cluster	Cluster
P1 (1,2)	C1 (1,2)	C2 (8,7)	C1
P2 (1,1)	0	5.6013	C1
P3 (2,2)	1	7.2193	C1
P4 (8,7)	8.6+23	7.5438	C2
P5 (9,8)	10	0	C2
P6 (8,8)	9.2194	1.9141	C2

① $C1(1,2)$ ② $C2(8,7)$
 $\frac{1}{\sqrt{(1-1)^2 + (2-2)^2}} = 1$

③ $(P3, C1)$
 $\sqrt{(2-1)^2 + (2-2)^2} = 1$

④ $(P4, C1)$
 $\sqrt{(8-1)^2 + (7-2)^2} = \sqrt{49+25}$

⑤ $(P5, C1)$
 $\sqrt{(9-1)^2 + (8-2)^2} = \sqrt{64+36}$

⑥ $(P6, C1)$
 $\sqrt{(8-1)^2 + (8-2)^2} = \sqrt{49+36}$

⑦ $(P1, C2)$
 $\sqrt{(1-8)^2 + (2-7)^2} = \sqrt{49+25}$

⑧ $(P2, C2)$
 $\sqrt{(1-8)^2 + (1-7)^2} = \sqrt{49+36}$

⑨ $(P3, C2)$
 $\sqrt{(2-8)^2 + (2-7)^2} = \sqrt{36+25}$

⑩ $(P4, C2)$

⑪ $(P5, C2)$
 $\sqrt{(9-8)^2 + (8-7)^2} = \sqrt{2}$

⑫ $(P6, C2)$
 $\sqrt{(8-8)^2 + (8-7)^2} = 1$

New Centroids:
 $C1 = P1, P2, P3$
 $C2 = P4, P5, P6$

$\Rightarrow \frac{(1+1+2)}{3} \Rightarrow \frac{1.33}{3} \Rightarrow \frac{2+1+2}{3} \Rightarrow 1.667$

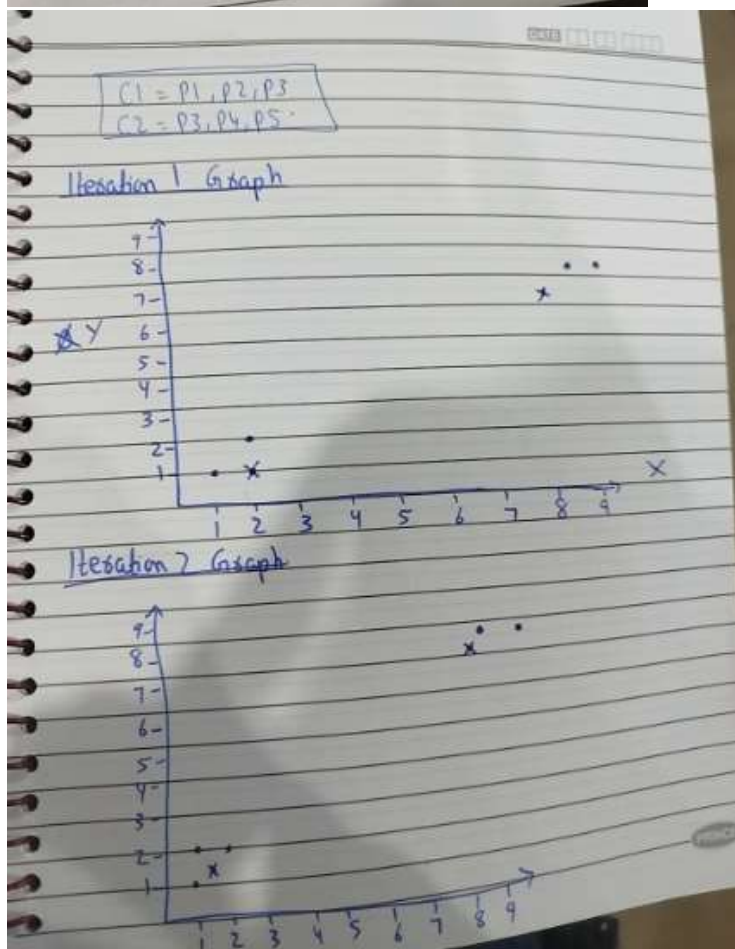
$C1 = (1.33, 1.667)$

$C2 = \frac{8+9+8}{3} \Rightarrow \frac{8.333}{3} \Rightarrow \frac{7+8+8}{3} \Rightarrow 7.667$

$C2 = (8.333, 7.667)$

Points	C1 (1.33, 1.66)	C2 (8.333, 7.666)	Cluster
P1 = (1,2)	0.1122	9.2676	C1
P2 = (1,1)	0.7379	9.9100	C1
P3 = (2,2)	0.7573	8.4977	C1
P4 = (8,7)	8.5443	0.7446	C2
P5 = (9,8)	9.1511	0.7460	C2
P6 = (8,8)	9.2024	0.4716	C2

- ① P_1, C_1
 $\sqrt{(1-1.33)^2 + (2-1.66)^2}$
- ② P_2, C_1
 $\sqrt{(1-1.33)^2 + (1-1.66)^2}$
- ③ P_3, C_1
 $\sqrt{(2-1.33)^2 + (2-1.66)^2}$
- ④ P_4, C_1
 $\sqrt{(8-1.33)^2 + (7-1.66)^2}$
- ⑤ P_5, C_1
 $\sqrt{(9-1.33)^2 + (8-1.66)^2}$
- ⑥ P_6, C_1
 $\sqrt{(8-1.33)^2 + (8-1.66)^2}$
- ⑦ P_1, C_2
 $\sqrt{(1-8.333)^2 + (2-7.666)^2}$
- ⑧ P_2, C_2
 $\sqrt{(1-8.333)^2 + (1-7.666)^2}$
- ⑨ P_3, C_2
 $= \sqrt{(2-8.333)^2 + (2-7.666)^2}$
- ⑩ P_4, C_2
 $\sqrt{(8-8.333)^2 + (7-7.666)^2}$
- ⑪ P_5, C_2
 $\sqrt{(9-8.333)^2 + (8-7.666)^2}$
- ⑫ P_6, C_2
 $\sqrt{(8-8.333)^2 + (8-7.666)^2}$



Code:

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

df = pd.DataFrame({
    "Point": ["P1", "P2", "P3", "P4", "P5", "P6"],
    "Feature1": [1, 1, 2, 8, 9, 8],
    "Feature3": [2, 1, 2, 7, 8, 8]
})

initial_centroids = df[["Feature1", "Feature3"]].iloc[[0, 3]].values

kmeans = KMeans(
    n_clusters=2,
    init=initial_centroids,
    n_init=1,
    random_state=42
)

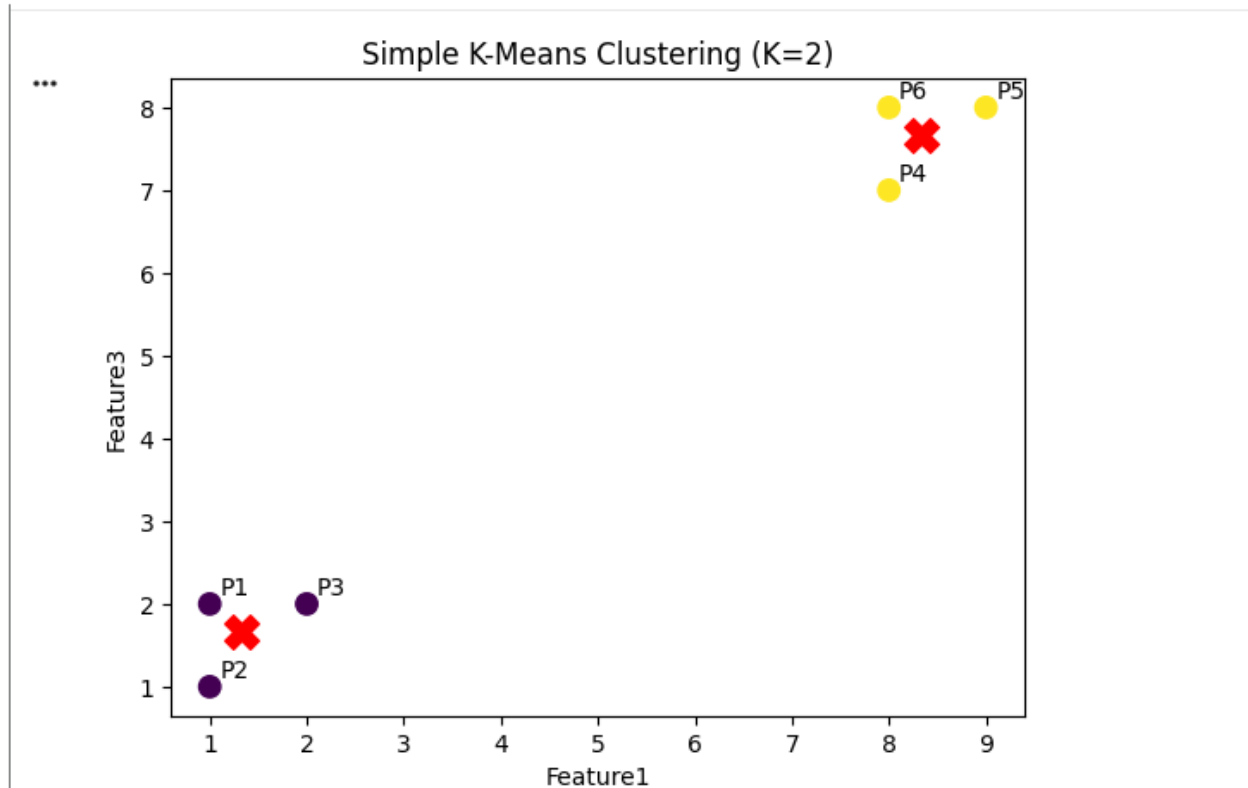
df["Cluster"] = kmeans.fit_predict(df[["Feature1", "Feature3"]])

plt.scatter(df["Feature1"], df["Feature3"],
            marker="o", s=80, c=df["Cluster"])

centroids = kmeans.cluster_centers_
plt.scatter(centroids[:, 0], centroids[:, 1],
            marker="X", s=200, c="red")

for i, txt in enumerate(df["Point"]):
    plt.annotate(txt, (df["Feature1"][i]+0.1, df["Feature3"][i]+0.1))

plt.xlabel("Feature1")
plt.ylabel("Feature3")
plt.title("Simple K-Means Clustering (K=2)")
plt.show()
```



Task2:

Scipy

```
import numpy as np
import pandas as pd
from scipy.cluster.vq import whiten, kmeans, vq
import matplotlib.pyplot as plt

data = pd.read_csv('Mall_Customers(in).csv')

features = data[['Annual Income (k$)', 'Spending Score (1-100)']].values

data_whitened = whiten(features)

k = 5
codebook, distortion = kmeans(data_whitened, k)

cluster_indices, _ = vq(data_whitened, codebook)

plt.figure(figsize=(10, 6))
plt.scatter(features[:, 0], features[:, 1], c=cluster_indices,
            cmap='viridis', s=50, alpha=0.8)
```

```
plt.title('K-Means Clustering of Mall Customers (K=5)')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.grid(True)
plt.show()
```

