Group 37 – Enzo Nunes 106336, Tomás Barros 106588

Machine Learning – Homework4

Ex1 – Clustering

i)

3.

$x^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad x^{(2)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \quad x^{(3)} = \begin{pmatrix} 0.5 \\ 0.55 \end{pmatrix}$

$p(x \mid c=1) = N\left(\mu^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$

$p(x \mid c=2) = N\left(\mu^2 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \Sigma^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$

$p(c=1) = 0,6 \quad p(c=2) = 0.4$

E-Step: Assign each point to the cluster that yields higher posterior

- $x^{(1)}$:
  - $c=1$
    - prior: $p(c=1) = 0.6$
    - likelihood $p(x^{(1)} \mid c=1) = \frac{1}{2\pi \det(\Sigma)} \exp\left(-\frac{1}{2}(x^{(1)}-\mu^1)^T (\Sigma)^{-1} (x^{(1)}-\mu^1)\right) =$

      $= \frac{1}{2\pi}$
    - joint probability: $p(c=1, x^{(1)}) = p(c=1)\, p(x^{(1)} \mid c=1) = 0,6 \times \frac{1}{2\pi} = \frac{0,6}{2\pi}$
  - $c=2$
    - prior: $p(c=2) = 0.4$
    - likelihood $p(x^{(1)} \mid c=2) = \frac{1}{2\pi \det(\Sigma)} \exp\left(-\frac{1}{2}(x^{(1)}-\mu^2)^T (\Sigma)^{-1} (x^{(1)}-\mu^2)\right) =$

      $= \frac{1}{2\pi} e^{-4}$
    - joint probability: $p(c=2, x^{(1)}) = p(c=2)\, p(x^{(1)} \mid c=2) = 0.4 \times \frac{1}{2\pi} e^{-4} = \frac{0.4}{2\pi} e^{-4}$
  - Normalizing the posteriors
    - $c=1$: $p(c=1 \mid x^{(1)}) = \frac{p(c=1, x^{(1)})}{p(c=1, x^{(1)}) + p(c=2, x^{(1)})} = \frac{\frac{0,6}{2\pi}}{\frac{0,6}{2\pi} + \frac{0,4}{2\pi} e^{-4}} \simeq 0.9879$
    - $c=2$ $p(c=2 \mid x^{(1)}) = \frac{p(c=2, x^{(1)})}{p(c=2, x^{(1)}) + p(c=1, x^{(1)})} = \frac{\frac{0,4}{2\pi} e^{-4}}{\frac{0,6}{2\pi} + \frac{0,4}{2\pi} e^{-4}} \simeq 0.0121$

- $x^{(2)}$:
  - $c=1$:
    - prior $p(c=1)=0.6$
    - likelihood $p(x^{(2)}|c=1)=\frac{1}{2\pi\det(\Sigma)}\exp\left(-\frac{1}{2}(x^{(2)}-\mu^1)^T(\Sigma)^{-1}(x^{(2)}-\mu^1)\right)=$
      $=\frac{1}{2\pi}e^{-4}$
    - joint probability: $p(c=1,x^{(2)})=p(c=1)p(x^{(2)}|c=1)=0.6\times\frac{1}{2\pi}e^{-4}=\frac{0.6}{2\pi}e^{-4}$
  - $c=2$
    - prior $p(c=2)=0.4$
    - likelihood $p(x^{(2)}|c=2)=\frac{1}{2\pi\det(\Sigma)}\exp\left(-\frac{1}{2}(x^{(2)}-\mu^2)^T(\Sigma)^{-1}(x^{(2)}-\mu^2)\right)=$
      $=\frac{1}{2\pi}$
    - joint probability: $p(c=2,x^{(2)})=p(c=2)p(x^{(2)}|c=2)=0.4\times\frac{1}{2\pi}=\frac{0.4}{2\pi}$
  - Normalizing the posteriors
    - $c=1$: $p(c=1|x^{(2)})=\frac{p(c=1,x^{(2)})}{p(c=1,x^{(2)})+p(c=2,x^{(2)})}=\frac{\frac{0.6}{2\pi}e^{-4}}{\frac{0.6}{2\pi}e^{-4}+\frac{0.4}{2\pi}}\simeq 0.0267$
    - $c=2$: $p(c=2|x^{(2)})=\frac{p(c=2,x^{(2)})}{p(c=2,x^{(2)})+p(c=1,x^{(2)})}=\frac{\frac{0.4}{2\pi}}{\frac{0.6}{2\pi}e^{-4}+\frac{0.4}{2\pi}}\simeq 0.9733$

- x^3
  - $c=1$
    - prior $p(c=1)=0.6$
    - likelihood $p(x^{(3)}|c=1)=\frac{1}{2\pi\det(\Sigma)}\exp\left(-\frac{1}{2}(x^{(3)}-\mu^1)^T(\Sigma)^{-1}(x^{(3)}-\mu^1)\right)=$
      $=\frac{1}{2\pi}e^{-0.22625}$
    - joint probability $p(c=1,x^{(3)})=p(c=1)p(x^{(3)}|c=1)=0.6\times\frac{1}{2\pi}e^{-0.22625}=\frac{0.6}{2\pi}e^{-0.22625}$
  - $c=2$
    - prior $p(c=2)=0.4$
    - likelihood $p(x^{(3)}|c=2)=\frac{1}{2\pi\det(\Sigma)}\exp\left(-\frac{1}{2}(x^{(3)}-\mu^2)^T(\Sigma)^{-1}(x^{(3)}-\mu^2)\right)=$
      $=\frac{1}{2\pi}e^{-2.32625}$
    - joint probability $p(c=2,x^{(3)})=p(c=2)p(x^{(3)}|c=2)=0.4\times\frac{1}{2\pi}e^{-2.32625}=\frac{0.4}{2\pi}e^{-2.32625}$
  - Normalizing the posteriors
    - $c=1$: $p(c=1|x^{(3)})=\frac{p(x^3|c=1)}{p(x^3|c=1)+p(x^3|c=2)}=\frac{\frac{0.6}{2\pi}e^{-0.22625}}{\frac{0.6}{2\pi}e^{-0.22625}+\frac{0.4}{2\pi}e^{-2.32625}}\simeq 0.9245$
    - $c=2$: $p(c=2|x^{(3)})=\frac{p(x^3|c=2)}{p(x^3|c=2)+p(x^3|c=1)}=\frac{\frac{0.4}{2\pi}e^{-2.32625}}{\frac{0.6}{2\pi}e^{-0.22625}+\frac{0.4}{2\pi}e^{-2.32625}}\simeq 0.0755$

M-Step: Re-estimate cluster parameters (new prior and likelihood)

$$\mu^c = \frac{\sum_{m=1}^{3} P(C=c \mid x^{(m)})\, x^{(m)}}{\sum_{m=1}^{3} P(C=c \mid x^{(m)})}$$

$$\Sigma_{ij}^c = \frac{\sum_{m=1}^{3} P(C=c \mid x^{(m)})(x_i^{(m)}-\mu_i^c)(x_j^{(m)}-\mu_j^c)}{\sum_{m=1}^{3} P(C=c \mid x^{(m)})}$$

$$P(C=c) = \frac{\sum_{m=1}^{N} P(C=c \mid x^{(m)})}{\sum_{\ell=1}^{k}\sum_{m=1}^{N} P(C=\ell \mid x^{(m)})}$$

- $C=1$
  - likelihood

$$-\mu^1 = \frac{0.9879\begin{pmatrix}1\\1\end{pmatrix}+0.0267\begin{pmatrix}-1\\-1\end{pmatrix}+0.9245\begin{pmatrix}0.5\\0.55\end{pmatrix}}{0.9879+0.0267+0.9245} \approx \begin{pmatrix}0.7341\\0.7579\end{pmatrix}$$

$$-\Sigma_{11}^1 = \frac{0.9879(1-0.7341)(1-0.7341)+0.0267(-1-0.7341)(-1-0.7341)+0.9245(0.5-0.3341)(0.5-0.7341)}{0.9879+0.0267+0.9245} = 0.1036$$

$$-\Sigma_{12}^1 = \Sigma_{21}^1 = \frac{0.9879(1-0.7341)(1-0.7579)+0.0267(-1-0.7341)(-1-0.7579)+0.9245(0.5-0.7341)(0.55-0.7579)}{0.9879+0.0267+0.9245} = 0.09797$$

$$-\Sigma_{22}^1 = \frac{0.9879(1-0.7579)(1-0.7579)+0.0267(-1-0.7579)(-1-0.7579)+0.9245(0.55-0.7579)(0.55-0.7579)}{0.9879+0.0267+0.9245} = 0.0930$$

$$-\Sigma^1 = \begin{pmatrix}0.1036 & 0.09797\\0.09797 & 0.0930\end{pmatrix} \quad P(x \mid C=1) = N\left(\mu^1 = \begin{pmatrix}0.7341\\0.7579\end{pmatrix}, \Sigma^1 = \begin{pmatrix}0.1036 & 0.09777\\0.09797 & 0.0930\end{pmatrix}\right)$$

- Prior

$$(C=1) = \frac{0.9879+0.0267+0.9245}{(0.9879+0.0267+0.9245)+(0.0121+0.9733+0.0755)} = 0.6464$$

- $C=2$
  - likelihood

$$-\mu^2 = \frac{0.0121\begin{pmatrix}1\\1\end{pmatrix} + 0.9733\begin{pmatrix}-1\\-1\end{pmatrix} + 0.0755\begin{pmatrix}0.5\\0.55\end{pmatrix}}{0.0121+0.9733+0.0755} \simeq \begin{pmatrix}-0.8704\\-0.8669\end{pmatrix}$$

$$-\Sigma^2_{11} = \frac{0.0121(1-(-0.8704))(1-(-0.8704)) + 0.9733(-1-(-0.8704))(-1-(-0.8704)) + 0.0755(0.5-(-0.8704))(0.5-(-0.8704))}{0.0121+0.9733+0.0755} =$$

$$= 0.1890$$

$$-\Sigma^2_{12} = \Sigma^2_{21} = \frac{0.0121(1-(-0.8704))(1-(-0.8669)) + 0.9733(-1-(-0.8704))(-1-(-0.8669)) + 0.0755(0.5-(-0.8704))(0.55-(-0.8669))}{0.0121+0.9733+0.0755} = 0.1938$$

$$-\Sigma^2_{22} = \frac{0.0121(1-(-0.8669))(1-(-0.8669)) + 0.9733(-1-(-0.8669))(-1-(-0.8669)) + 0.0755(0.55-(-0.8669))(0.55-(-0.8669))}{0.0121+0.9733+0.0755} = 0.1989$$

$$-\Sigma^2 = \begin{pmatrix}0.1890 & 0.1938\\0.1938 & 0.1989\end{pmatrix} \quad p(x \mid c=2) = N\left(\mu^2 = \begin{pmatrix}-0.8704\\-0.8669\end{pmatrix}, \Sigma^2 = \begin{pmatrix}0.1890 & 0.1938\\0.1938 & 0.1989\end{pmatrix}\right)$$

- Prior

$$P(C=2) = \frac{0.0121+0.9733+0.0755}{(0.0121+0.9733+0.0755)+(0.9879+0.0267+0.9245)} = 0.3536$$

ii)

```python
from sklearn.metrics import silhouette_samples
import numpy as np

# From the previous exercise
X = np.array([[1, 1], [-1, -1], [0.5, 0.55]])
cluster_probabilities = np.array([[0.9879, 0.0121], [0.0267, 0.9733], [0.9245, 0.0755]])

# Perform the hard assignment
hard_assignments = np.argmax(cluster_probabilities, axis=1)

# Identify the larger cluster
unique, counts = np.unique(hard_assignments, return_counts=True)
larger_cluster = unique[np.argmax(counts)]

# Calculate silhouette values for the larger cluster
silhouette_vals = silhouette_samples(X, hard_assignments)

# Filter out the silhouette values for the larger cluster
silhouette_larger_cluster = silhouette_vals[hard_assignments == larger_cluster]

# Calculate the average silhouette score for the larger cluster
average_silhouette = np.mean(silhouette_larger_cluster)
```

The silhouette of the larger cluster is 0.7252.

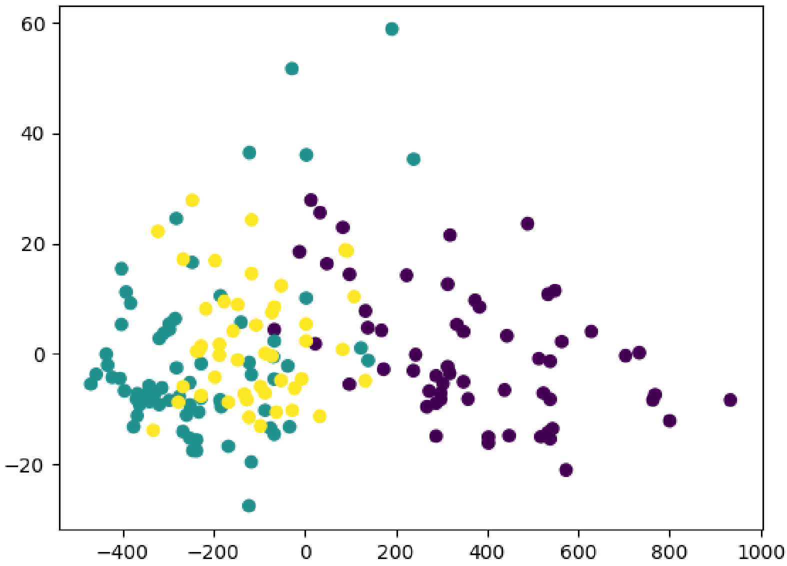
Ex2 – Software Experiments

a)

With the *Wine* dataset, the silhouette for the *k-Means* algorithm is 0.5711, and for *EM Clustering* is 0.2833. This means the *k-Means* created more well-separated and cohesive clusters. Thus, in this case, *k-Means* is a better algorithm.

b)

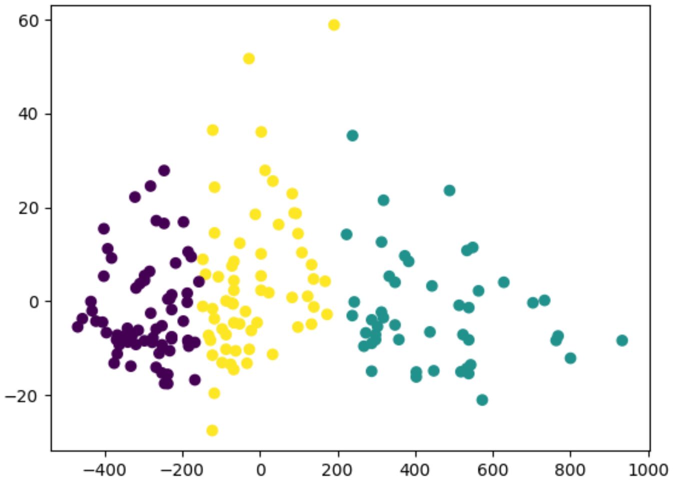
With the *Wine* dataset, the *PCA* with 2 components yields the following scatter plot:



The three classes cannot be fully separated using *PCA* with two components due to significant overlap between two of the groups, which can be noticed in the scatter plot.

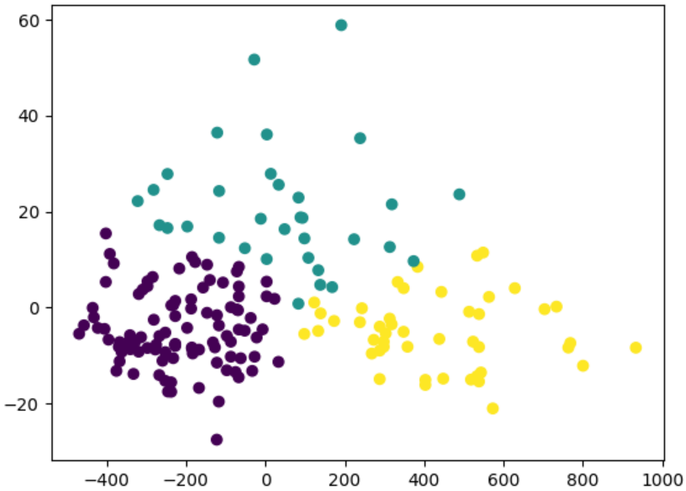
c)

With the dataset from question **b)**, the *k-Means* algorithm yields the following scatter plot:



Silhouette: 0.5723

And the *Em Clustering* yields:



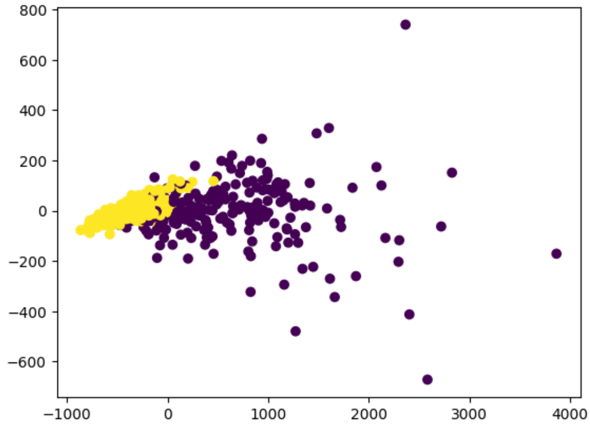
Silhouette: 0.3352

- The silhouette values here are different because the dataset used in this exercise has undergone dimensionality reduction from the *PCA*, which might have altered the relationships between data points, which can affect the effectiveness of clustering.
- The *EM clustering* approach seems to have produced clusters with more ambiguous boundaries, indicating that the *k-Means* was more effective in distinguishing these groups compared to *EM Clustering*.

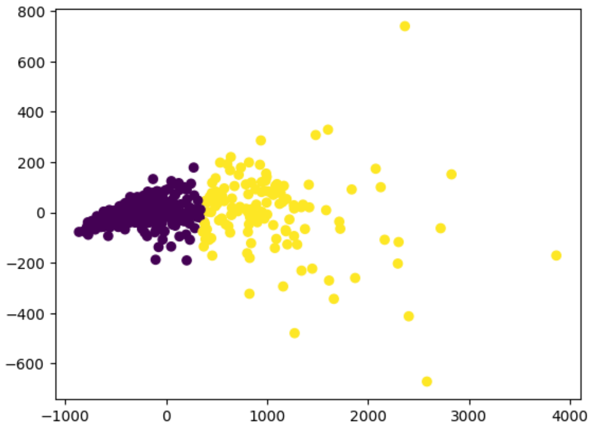
d)

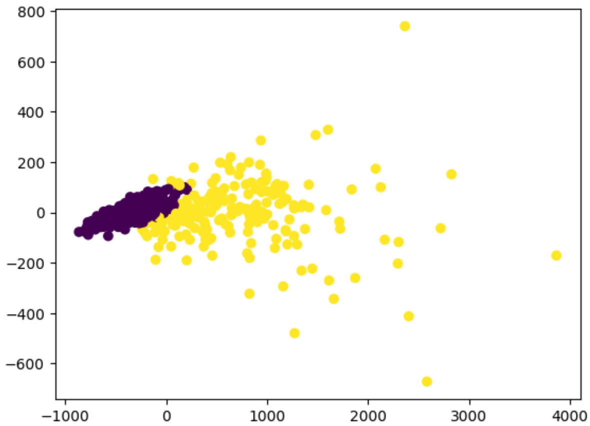
Repeating the previous tests but with the breast cancer dataset and 2 clusters for each algorithm, the results are the following:

- Without PCA, the silhouettes for the *k-Means* and *EM Clustering* are, respectively, 0.6973 and 0.5315. The scatter plot for the *PCA*:



- With the *PCA*-mapped data, the silhouettes for the *k-Means* and *EM Clustering* are, respectively, 0.6984 and 0.5865. The scatter plots are, respectively:





- Again, the *k-Means* shows a better separation of the clusters and has a higher silhouette score. This indicates that it was more effective in distinguishing both data groups compared to the *EM Clustering*.