TÉCNICO LISBOA

Aprendizagem 2022/23

Homework IV – Group 102

II. Pen and paper

1) $x_1 = \begin{bmatrix}1\\2\end{bmatrix}$ $x_2 = \begin{bmatrix}-1\\1\end{bmatrix}$ $x_3 = \begin{bmatrix}1\\0\end{bmatrix}$

$$-\frac{1}{2}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}\mu_1\\\mu_2\end{bmatrix}\right)^T \Sigma^{-1}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}\mu_1\\\mu_2\end{bmatrix}\right)$$

Para $x_1 = -\frac{1}{2}\begin{bmatrix}-1\\0\end{bmatrix}^T \begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}^{-1}\begin{bmatrix}-1\\0\end{bmatrix}$

$$= \text{—} \cdot \begin{bmatrix}2/3 & -1/3\\-1/3 & 2/3\end{bmatrix}$$

$$= -1/3$$

$x_2 = -\frac{1}{2}[-3,-1]\,\Sigma^{-1}\begin{bmatrix}-3\\-1\end{bmatrix} =$

$$= \frac{-7}{3}$$

$x_3 = -\frac{1}{2}[-1,-2]\,\Sigma^{-1}\begin{bmatrix}-1\\-2\end{bmatrix}$

$$= -1$$

$\Sigma_2 \to x_1 = -\frac{1}{2}[1\ 2]\,\Sigma^{-1}\begin{bmatrix}1\\2\end{bmatrix}$

$$= -\frac{5}{4}$$

$x_2 = -\frac{1}{2}[-1\ 1]\,\Sigma^{-1}\begin{bmatrix}-1\\1\end{bmatrix}$

$$= -\frac{1}{2}$$

$x_3 = -\frac{1}{2}[1\ 0]\,\Sigma^{-1}\begin{bmatrix}1\\0\end{bmatrix}$

$$= -\frac{1}{4}$$

$$N\left(\begin{bmatrix}x_1\\x_2\end{bmatrix} \middle| \begin{bmatrix}\mu_1\\\mu_2\end{bmatrix}, \Sigma\right) = \frac{1}{(2\pi)^{\frac{2}{2}}} \times \frac{1}{\sqrt{\det(\Sigma)}} \times e^{-}$$

$\det(\Sigma_1) = 3$

$\det(\Sigma_2) = 4$

$$\Sigma_1 \rightarrow x_1 \Rightarrow \frac{1}{2\pi} \times \frac{1}{\sqrt{\det(\Sigma_1)}} \times e^{-\frac{1}{3}}$$
$$= 0{,}066$$

$$x_2 \Rightarrow \frac{1}{2\pi} \times \frac{1}{\sqrt{\det(\Sigma_1)}} \times e^{-\frac{7}{3}}$$
$$= 0{,}009$$

$$x_3 \Rightarrow \frac{1}{2\pi} \times \frac{1}{\sqrt{\det(\Sigma_1)}} \times e^{-1}$$
$$= 0{,}034$$

$$\Sigma_2 \Rightarrow x_1 \Rightarrow \frac{1}{2\pi} \times \frac{1}{2} \times e^{-\frac{5}{4}}$$
$$= 0{,}023$$

$$x_2 \Rightarrow \frac{1}{2\pi} \times \frac{1}{2} \times e^{-\frac{1}{2}}$$
$$= 0{,}048$$

$$x_3 \Rightarrow \frac{1}{2\pi} \times \frac{1}{2} \times e^{-\frac{1}{4}}$$
$$= 0{,}062$$

$$P(x_1) = 0{,}5 \times 0{,}066 + 0{,}5 \times 0{,}023 = 0{,}045$$

$$P(x_2) = 0{,}5 \times 0{,}009 + 0{,}5 \times 0{,}048 = 0{,}029$$

$$P(x_3) = 0{,}5 \times 0{,}034 + 0{,}5 \times 0{,}062 = 0{,}048$$

$$P(k=1 \mid x_1) = \frac{0{,}5 \times 0{,}066}{0{,}045} = 0{,}733$$

$$P(k=2 \mid x_1) = \frac{0{,}5 \times 0{,}023}{0{,}045} = 0{,}256$$

$$P(k=1 \mid x_2) = \frac{0{,}5 \times 0{,}009}{0{,}029} = 0{,}155$$

$$P(k=2 \mid x_2) = \frac{0{,}5 \times 0{,}048}{0{,}029} = 0{,}828$$

$$P(k=1 \mid x_3) = \frac{0{,}5 \times 0{,}034}{0{,}048} = 0{,}354$$

$$P(k=2 \mid x_3) = \frac{0{,}5 \times 0{,}062}{0{,}048} = 0{,}646$$

$$u_1 = \frac{0{,}733 \times \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0{,}155 \begin{bmatrix} -1 \\ 1 \end{bmatrix} + 0{,}354 \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{0{,}733 + 0{,}155 + 0{,}354}$$

$$= \frac{\begin{bmatrix} 0{,}733 \\ 1466 \end{bmatrix} + \begin{bmatrix} -0{,}155 \\ 0{,}155 \end{bmatrix} + \begin{bmatrix} 0{,}354 \\ 0 \end{bmatrix}}{1{,}242}$$

$$= \begin{bmatrix} 0{,}750 \\ 1{,}305 \end{bmatrix}$$

$$u_2 = \frac{0{,}256 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0{,}828 \begin{bmatrix} -1 \\ 1 \end{bmatrix} + 0{,}646 \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{0{,}256 + 0{,}828 + 0{,}646}$$

$$= \begin{bmatrix} 0{,}043 \\ 0{,}775 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} \sigma_1^{11} & \sigma_1^{12} \\ \sigma_1^{21} & \sigma_1^{22} \end{bmatrix}$$

$$\sigma_1^{11} = \frac{0{,}733(1-0{,}750)^2 + 0{,}155(-1-0{,}75)^2 + 0{,}354(1-0{,}75)^2}{0{,}733 + 0{,}155 + 0{,}354}$$

$$= 0{,}437$$

$$\sigma_1^{22} = 0{,}782$$

$$\sigma_1^{12} = \sigma_1^{21} = 0{,}076$$

$$\Sigma_2 = \begin{bmatrix} \sigma_2^{11} & \sigma_2^{12} \\ \sigma_2^{21} & \sigma_2^{22} \end{bmatrix}$$

$$\sigma_2^{11} = \frac{0{,}256(1-0{,}043)^2 + 0{,}828(-1-0{,}043)^2 + 0{,}646(1-0{,}043)}{0{,}256 + 0{,}828 + 0{,}646}$$

$$= 0{,}998$$

$$\sigma_2^{22} = 0{,}471$$

$$\sigma_2^{12} = \sigma_2^{21} = -0{,}216$$

$$\pi_1 = \frac{0{,}733 + 0{,}155 + 0{,}354}{3} = 0{,}414$$

$$\pi_2 = 1 - 0{,}414 = 0{,}586$$

2) $\Sigma_1 = \begin{bmatrix} 0{,}437 & 0{,}076 \\ 0{,}076 & 0{,}782 \end{bmatrix}$ $\Sigma_2 = \begin{bmatrix} 0{,}998 & -0{,}216 \\ -0{,}216 & 0{,}471 \end{bmatrix}$

$\mu_1 = \begin{bmatrix} 0{,}750 \\ 1{,}305 \end{bmatrix}$ $\mu_2 = \begin{bmatrix} 0{,}043 \\ 0{,}775 \end{bmatrix}$

Σ_1:

x_1: $-\frac{1}{2}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 0{,}750 \\ 1{,}305 \end{bmatrix}\right)^T \Sigma_1^{-1}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 0{,}750 \\ 1{,}305 \end{bmatrix}\right)$

$= -\frac{1}{2} \times 0{,}695$

$= -0{,}348$

x_2: $-\frac{1}{2}\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0{,}750 \\ 1{,}305 \end{bmatrix}\right)^T \Sigma_1\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \mu_1\right)$

$= -3{,}5045$

$x_3 = -1{,}2545$

Σ_2:

x_1: $-2{,}876$

x_2: $-0{,}545$

x_3: $-0{,}839$

$\det(\Sigma_1) = 0,336 \quad \det(\Sigma_2) = 0,423$

$P(x_1 \mid k=1) = \frac{1}{2\pi} \times \frac{1}{\sqrt{0,336}} \times e^{-0,348}$

$= 0,194$

$P(x_2 \mid k=1) = 0,008$

$P(x_3 \mid k=1) = 0,078$

$P(x_1 \mid k=2) = \frac{1}{2\pi} \times \frac{1}{\sqrt{0,423}} \times e^{-2,876} = 0,014$

$P(x_2 \mid k=2) = 0,142$

$P(x_3 \mid k=2) = 0,106$

$P(x_1) = P(k=1)P(x_1 \mid k=1) + P(k=2)P(x_1 \mid k=2)$

$= 0,089$

$P(x_2) = 0,087$

$P(x_3) = 0,094$

$P(k=1 \mid x_1) = \frac{P(k=1)P(x_1 \mid k=1)}{P(x_1)}$

$= 0,902$

$P(k=1 \mid x_2) = 0,038$

$P(k=1 \mid x_3) = 0,\cancel{0}343$

$x_1 \in k_1$ porque $P(k_1=1 \mid x_1) > 0,5 > P(k=2 \mid x_1)$

$x_2 \in k_2$ porque $P(k_1=1 \mid x_2) < 0,5 < P(k=2 \mid x_2)$

$x_3 \in k_2$ porque $P(k_1=1 \mid x_3) < 0,5 < P(k=2 \mid x_3)$

2) b) $S(k=2) = \frac{S(x_2) + S(x_3)}{2}$ $\quad S(x_m) = 1 - \frac{a(x_m)}{b(x_m)}$

	x_1	x_2	x_3
x_1	0	$\sqrt{5}$	2
x_2	$\sqrt{5}$	0	$\sqrt{5}$
x_3	2	$\sqrt{5}$	0

$a(x_2) = \sqrt{5}$, $b(x_2) = \sqrt{5}$ $\Rightarrow S(x_2) = 1 - \frac{\sqrt{5}}{\sqrt{5}} = 0$

$a(x_3) = \sqrt{5}$, $b(x_3) = 2$ $\Rightarrow S(x_3) = 1 - \frac{\sqrt{5}}{2} = -0{,}118$

$S(k=2) = \frac{-0{,}118}{2} = -0{,}059$

Aprendizagem 2022/23

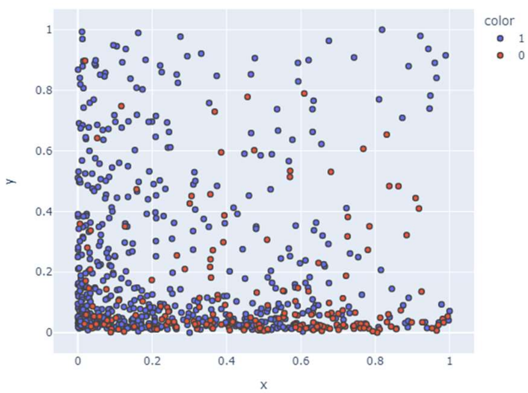
Homework IV – Group 102

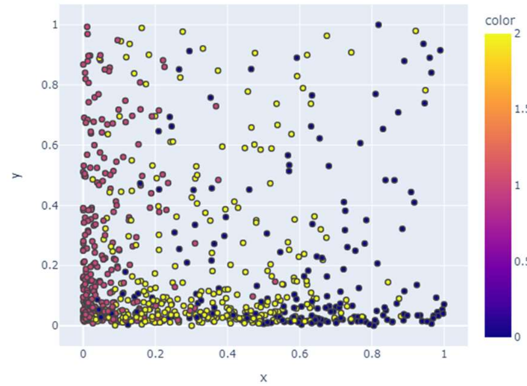
II. Programming and critical analysis

1. k=0 silhouette score : 0.11362027575179431
   k=1 silhouette score : 0.11403554201377074
   k=2 silhouette score : 0.11362027575179431

   k=0 K-means purity score : 0.7671957671957672
   k=1 K-means purity score : 0.7632275132275133
   k=2 K-means purity score : 0.7671957671957672

2.





3. Since the algorithm doesn't guarantee convergence to the global optimum, the result may depend on the initial clusters. Running it with different random states will end in different local maximums.

4. Needed 31 principal components to reach Explained variance = 0.8006422402169664

III. APPENDIX

```
import pandas as pd, numpy as np
import matplotlib.pyplot as plt
from scipy.io.arff import loadarff

from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
from sklearn.metrics.cluster import contingency_matrix
import plotly.express as px
from sklearn.decomposition import PCA

# import dataset
data = loadarff('drive/MyDrive/ML/pd_speech.arff')
df = pd.DataFrame(data[0])
df['class'] = df['class'].str.decode('utf-8')
X = df.drop('class', axis=1)
y = df['class']

# normalize the data
X_old = X
```

```
scaler = MinMaxScaler()
X = scaler.fit_transform(X)

# calculate pirity score
def purity_score(y_pred, y):
  confusion_matrix = contingency_matrix(y, y_pred)
  return np.sum(np.amax(confusion_matrix, axis=0)) / np.sum(confusion_matrix)

# predict k-means with random_state = {0,1,2}
kmeans_predict = []
for i in range(3):
  kmeans = KMeans(n_clusters=3, random_state=i).fit(X)
  kmeans_predict.append(kmeans.predict(X))
  print('k=' + str(i) +" silhouette score \t:", silhouette_score(X, kmeans_predict[i]))
for i in range(3):
  print('k=' + str(i) +" K-means purity score:", purity_score(kmeans_predict[i], y))

# Select the two input variables with highest variance
variances = np.var(X, axis=0)
idx = np.argsort(variances)[::-1]
X_new = X[:,idx[:2]]

# flip X so its a [700,2]
test = np.empty((2, 0)).tolist()
for i in X_new:
  test[0]. append(i[0])
  test[1]. append(i[1])

# print scatter from the original Parkinson diagnoses
fig = px.scatter(x = test[0], y = test[1], color = y)
fig.update_traces(marker_line_width=1.5)
fig.show()

# print scatter from k-means with random_state=0
fig = px.scatter(x = test[0], y = test[1], color = kmeans_predict[0])
fig.update_traces(marker_line_width=1.5)
fig.show()

# find number of components needed to explain more than 80% of variability
for i in range(len(y)):
  pca = PCA(n_components=i, svd_solver='full')
  pca.fit(X)
  var = 0
  for j in pca.explained_variance_ratio_:
    var += j
  if var >= 0.8:
    print( 'Needed', i, 'principal components to reach Explained variance =', var)
    break
```

END