TÉCNICO LISBOA

Aprendizagem 2022/23

Homework IV

Deadline 7/11/2022 (Monday) 23:59 via Fenix as PDF

1)

E - Step

x_1

$$P(K \mid x_1) = \frac{P(x_1 \mid K)\,P(K)}{P(x_1)}$$

K = 1

$$P(x_1 \mid K=1) \sim N\left(\mu_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right)$$

$$= \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(x_1-\mu_1)^T \Sigma_1^{-1} (x_1-\mu_1)}$$

$$|\Sigma_1| = 2\times 2 - 1\times 1 = 3$$

$$\Sigma_1^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}$$

$$x_1 - \mu_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$= \frac{1}{2\pi\sqrt{3}} e^{-\frac{1}{2}\begin{bmatrix} -1 & 0 \end{bmatrix}\begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}\begin{bmatrix} -1 \\ 0 \end{bmatrix}}$$

$$= \frac{1}{2\pi\sqrt{3}} e^{-1/3}$$

$$\simeq 0.0658$$

$$\text{Posterior}_1 = P(x_1 \mid K=1) \times \text{Prior}_1$$

$$= 0.1317 \times 0.5$$

$$= 0.033$$

$K = 2$

$$P(x_1 \mid K=2) \sim N\left(\mu_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)$$

$$= \frac{1}{2\pi \sqrt{|\Sigma_2|}} e^{-\frac{1}{2}(x_1 - \mu_2)^T \Sigma_2^{-1} (x_1 - \mu_2)}$$

$$|\Sigma_2| = 2 \times 2 = 4$$

$$\Sigma_2^{-1} = \frac{1}{4}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}$$

$$x_1 - \mu_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$= \frac{1}{2\pi \sqrt{4}} e^{-\frac{1}{2}\begin{bmatrix} 1 & 2 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix}}$$

$$= \frac{1}{2\pi \sqrt{4}} e^{-5/4}$$

$$\simeq 0.0228$$

$$\text{Posterior}_2 = P(x_1 \mid K=2) \times \text{Prior}_2$$

$$= 0.0228 \times 0.5$$

$$= 0.0114$$

$$P(K=1 \mid x_1) = \frac{0.033}{0.033 + 0.0114}$$

$$= 0.7434$$

$$P(K=2 \mid x_1) = \frac{0.0114}{0.033 + 0.0114}$$

$$= 0.2568$$

$\underset{\sim}{x}_2$

$$P(k \mid x_2) = \frac{P(x_2 \mid k)\, P(k)}{P(x_2)}$$

$k=1$

$$P(x_2 \mid k=1) \sim N\left(\mu_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right)$$

$$= \frac{1}{2\pi \sqrt{|\Sigma|}} e^{-\frac{1}{2}(x_2-\mu_1)^T \Sigma_1^{-1} (x_2-\mu_1)}$$

$$|\Sigma_1| = 2\times 2 - 1\times 1 = 3$$

$$\Sigma_1^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}$$

$$x_2 - \mu_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -3 \\ -1 \end{bmatrix}$$

$$= \frac{1}{2\pi\sqrt{3}} e^{-\frac{1}{2}\begin{bmatrix} -3 & -1 \end{bmatrix}\begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}\begin{bmatrix} -3 \\ -1 \end{bmatrix}}$$

$$= \frac{1}{2\pi\sqrt{3}} e^{-7/3}$$

$$\simeq 0.00891$$

$$\text{Posterior}_1 = P(x \mid k=1) \times \text{Prior}_1$$

$$= 9.7716 \times 0.5$$

$$= 0.00446$$

$K = 2$

$$P(x_2 \mid K=2) \sim N\left(\mu_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)$$

$$= \frac{1}{2\pi \sqrt{|\Sigma_2|}} e^{-\frac{1}{2}(x_2 - \mu_2)^T \Sigma_2^{-1} (x_2 - \mu_2)}$$

$$|\Sigma_2| = 2 \times 2 = 4$$

$$\Sigma_2^{-1} = \frac{1}{4}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}$$

$$x_2 - \mu_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$= \frac{1}{2\pi\sqrt{4}} e^{-\frac{1}{2}\begin{bmatrix} -1 & 1 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix}}$$

$$= \frac{1}{2\pi\sqrt{4}} e^{-1/2}$$

$$\simeq 0.0483$$

$$\text{Posterior}_2 = P(x \mid K=2) \times \text{Prior}_2$$
$$= 0.0483 \times 0.5$$
$$= 0.0241$$

$$P(K=1 \mid x_2) = \frac{0.00446}{0.0241 + 0.00446}$$

$$= 0.1562$$

$$P(K=2 \mid x_2) = \frac{0.0241}{0.0241 + 0.00446}$$

$$= 0.8438$$

$\underset{\sim}{x}_3$

$$P(K \mid x_3) = \frac{P(x_3 \mid K)\,P(K)}{P(x_3)}$$

$K = 1$

$$P(x_3 \mid K=1) \sim N\left(\mu_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right)$$

$$= \frac{1}{2\pi \sqrt{|\Sigma_1|}} e^{-\frac{1}{2}(x_3 - \mu_1)^T \Sigma_1^{-1} (x_3 - \mu_1)}$$

$$|\Sigma_1| = 2 \times 2 - 1 \times 1 = 3$$

$$\Sigma_1^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}$$

$$x_3 - \mu_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

$$= \frac{1}{2\pi \sqrt{3}} e^{-\frac{1}{2}\begin{bmatrix} -1 & -2 \end{bmatrix}\begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}\begin{bmatrix} -1 \\ -2 \end{bmatrix}}$$

$$= \frac{1}{2\pi \sqrt{3}} e^{-1}$$

$$\simeq 0.0338$$

$$\text{Posterior}_1 = P(x \mid K=1) \times \text{Prior}_1$$
$$= 0.0338 \times 0.5$$
$$= 0.0169$$

$\underline{K=2}$

$$P(x_3 \mid K=2) \sim N\left(\mu_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)$$

$$= \frac{1}{2\pi \sqrt{|\Sigma_2|}} e^{-\frac{1}{2}(x_3-\mu_2)^T \Sigma_2^{-1} (x_3-\mu_2)}$$

$$|\Sigma_2| = 2 \times 2 = 4$$

$$\Sigma_2^{-1} = \frac{1}{4}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}$$

$$x_3 - \mu_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= \frac{1}{2\pi \sqrt{4}} e^{-\frac{1}{2}\begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}}$$

$$= \frac{1}{2\pi \sqrt{4}} e^{-1/4}$$

$$\simeq 0.062$$

$$\text{Posterior}_2 = P(x \mid K=2) \times \text{Prior}_2$$
$$= 0.062 \times 0.5$$
$$= 0.031$$

$$P(K=1 \mid x_3) = \frac{0.0169}{0.031 + 0.0169}$$

$$= 0.3528$$

$$P(K=2 \mid x_3) = \frac{0.031}{0.031 + 0.0169}$$

$$= 0.6472$$

M - Step

$$\mu_c = \frac{\sum_{i=1}^{N} P(c|x_i) \cdot x_i}{\sum_{i=1}^{N} P(c|x_i)}$$

$$Prior_c = p(c = k) = \frac{\sum P(c \mid x_i)}{\underbrace{\sum_{j=1}^{k} P(C=j|x_i)}_{=N}}$$

$$\Sigma_c^{ij} = \frac{\sum_{k=1}^{N} p(c|k) \cdot (x_k^i - \mu_c^i)(x_k^j - \mu_c^j)}{\sum_{k=1}^{N} p(c|x_k)}$$

$c = 1$

$$\sum_{i=1}^{N} p(c|x_i) = 0.7434 + 0.1562 + 0.3538$$
$$= 1.2534$$

$$\mu_1 = \frac{0.7434 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0.1562 \begin{bmatrix} -1 \\ 1 \end{bmatrix} + 0.3538 \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{1.2534}$$

$$= \begin{bmatrix} 0.7508 \\ 1.3108 \end{bmatrix}$$

$$\Sigma_1^{(1,1)} = \frac{0.7434(1 - 0.7508)^2 + 0.1562(-1 - 0.7508)^2 + 0.3538(1 - 0.7538)^2}{1.2534}$$

$$= 0.4359$$

$$\Sigma^{(2,1)} = \Sigma^{(1,2)}$$

$$= \frac{0.7434(1 - 0.7508)(2 - 1.3108) + 0.1562(-1 - 0.7508)(1 - 1.3108) + 0.3538(1 - 0.7508)(0 - 1.3108)}{1.2534}$$

$$= 0.0775$$

$$\Sigma_1^{(2,2)} = \frac{0.7434(2-1.3108)^2 + 0.1562(1-1.3108)^2 + 0.3538(0-1.3108)^2}{1.2534}$$

$$= 0.7788$$

Therefore, $\Sigma_1 = \begin{bmatrix} 0.4359 & 0.0775 \\ 0.0775 & 0.7788 \end{bmatrix}$

$$\pi_1 = \frac{1.2534}{3}$$

$$= 0.4178$$

$\underline{c = 2}$

$$\sum_{i=1}^{N} P(c|x_i) = 0.2568 + 0.8438 + 0.6472$$
$$= 1.7478$$

$$\mu_2 = \frac{0.2568\begin{bmatrix}1\\2\end{bmatrix} + 0.8438\begin{bmatrix}-1\\1\end{bmatrix} + 0.6472\begin{bmatrix}1\\0\end{bmatrix}}{1.7478}$$

$$= \begin{bmatrix} 0.0344 \\ 0.7766 \end{bmatrix}$$

$$\Sigma_2^{(1,1)} = \frac{0.2568(1-0.0344)^2 + 0.8438(-1-0.0344)^2 + 0.6472(1-0.0344)^2}{1.7478}$$

$$= 0.9988$$

$$\Sigma^{(2,1)} = \Sigma^{(1,2)}$$

$$= \frac{0.2568(1-0.0344)(2-0.7766) + 0.8438(-1-0.0344)(1-0.7766) + 0.6472(1-0.0344)(0-0.7766)}{1.7478}$$

$$= -0.2157$$

$$\Sigma^{(2,2)} = \frac{0.2568(2-0.7766)^2 + 0.8438(1-0.7766)^2 + 0.6472(0-0.7766)^2}{1.7478}$$

$$= 0.6517$$

Therefore, $\Sigma_2 = \begin{bmatrix} 0.9988 & -0.2157 \\ -0.2157 & 0.4673 \end{bmatrix}$

$$\tilde{\pi}_2 = \frac{1.7478}{3}$$

$$= 0.5826$$

To summarize, the new parameters are:

$$\mu_1 = \begin{bmatrix} 0.7508 \\ 1.3108 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} 0.0344 \\ 0.7766 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} 0.4359 & 0.0775 \\ 0.0775 & 0.7788 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 0.9988 & -0.2157 \\ -0.2157 & 0.4673 \end{bmatrix}$$

$$\tilde{\pi}_1 = 0.4178 \qquad \tilde{\pi}_2 = 0.5826 \quad //$$

2.a) Since we are considering a hard assignment under MAP assumption, we must assign each observation to the highest posterior probability:

$$P(k=1 \mid x_1) = \frac{0.033}{0.033 + 0.0114} = 0.7434$$

$$P(k=1 \mid x_2) = \frac{0.00446}{0.0241 + 0.00446} = 0.1562$$

$$P(k=1 \mid x_3) = \frac{0.0169}{0.031 + 0.0169} = 0.3528$$

$$P(k=2 \mid x_1) = \frac{0.0114}{0.033 + 0.0114} = 0.2568$$

$$P(k=2 \mid x_2) = \frac{0.0241}{0.0241 + 0.00446} = 0.8438$$

$$P(k=2 \mid x_3) = \frac{0.031}{0.031 + 0.0169} = 0.6472$$

Therefore, x_1 is assigned to cluster 1 and x_2 and x_3 are assigned to cluster 2. //

2.b) Silhouette of C_2 (largest cluster) using Euclidean distance:

$$S(C_2) = \frac{S(x_2) + S(x_3)}{2}$$

$a(x_2) = d(x_2, x_3)$
$= \sqrt{(-1-1)^2 + (1-0)^2}$
$= \sqrt{5}$

$b(x_2) = d(x_2, x_1)$
$= \sqrt{(-1-1)^2 + (1-2)^2}$
$= \sqrt{5}$

$a(x_2) \leq b(x_2)$
$\Downarrow$
$S(x_2) = 1 - \frac{a(x_2)}{b(x_2)}$

$= 1 - 1$

$= 0$

$a(x_3) = d(x_3, x_2)$
$= a(x_2)$
$= \sqrt{5}$

$b(x_3) = d(x_3, x_1)$
$= \sqrt{(1-1)^2 + (0-2)^2}$
$= 2$

$a(x_3) > b(x_3)$
$\Downarrow$
$S(x_3) = \frac{b(x_3)}{a(x_3)} - 1$

$= \frac{2}{\sqrt{5}} - 1$

$\simeq -0.1056$

$$= \frac{0 - 0.1056}{2}$$

$$= -0.0528$$

Therefore, the silhouette of the largest cluster is $S(C_2) = -0.0528$ //

II. Programming and critical analysis

The code for the questions is in the Appendix of this document.

1) Seed: 0

   Silhouette: 0.11362027575179438

   Purity: 0.7671957671957672

   Seed: 1

   Silhouette: 0.1140355420137708

   Purity: 0.7632275132275133

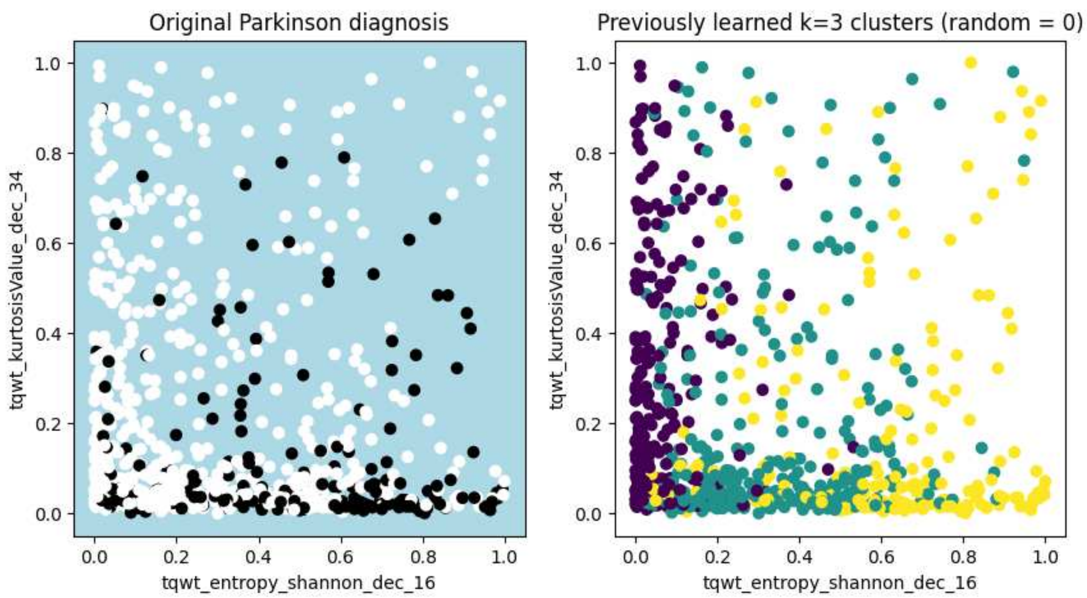
   Seed: 2

   Silhouette: 0.11362027575179438

   Purity: 0.7671957671957672

2) In the previous question, we observed that the results of the silhouette and purity were non-deterministic.
   The **`cluster.KMeans()`** function has a parameter called **`random_state`**. This parameter determines random number generation for centroid initialization and we use an integer in this parameter to make the randomness deterministic.
   Different **`random_state`** values will not prevent the algorithm from converging to the same final point since we can see that seed 0 and seed 2 results are the same.
   However, the results for seed 1 are different from the others due to the randomness in the initialization of the centroids and that is what is causing the non-determinism.

3)



4) Number of components: 31

III. APPENDIX

Code used in question 1) of **II. Programming and critical analysis**:

Loads the data:

```
import numpy as np
import pandas as pd
from scipy.io.arff import loadarff
from sklearn.metrics import silhouette_score

# Load the data
data = loadarff('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']
```

Calculates the silhouettes and purities:

```
from sklearn import datasets, metrics, cluster, mixture
from sklearn.preprocessing import MinMaxScaler

#normalize the data with MinMaxScaler
scaler = MinMaxScaler()
X = scaler.fit_transform(X)
X = pd.DataFrame(X, columns=df.columns[:-1])

#compute purity
def purity_score(y_true, y_pred):
   # compute contingency/confusion matrix
   confusion_matrix = metrics.cluster.contingency_matrix(y_true, y_pred)
   return np.sum(np.amax(confusion_matrix, axis=0)) / np.sum(confusion_matrix)

seeds = [0, 1, 2]
#apply k-means clustering fully unsupervised (without targets) on the normalized data with
k = 3 and three different seeds (using random ε {0,1,2}).
for seed in seeds:
   #parameterize clustering
   kmeans = cluster.KMeans(n_clusters=3, random_state=seed)

   #learn model
   kmeans_model = kmeans.fit(X)

    # check the produced clusters
    y_pred = kmeans_model.labels_
```

```
    print("Seed: ", seed)
    print("Silhouette:", metrics.silhouette_score(X, y_pred, metric='euclidean'))
    print("Purity:", purity_score(Y, y_pred))
    print("\n")
```

Code used in question 3) of **II. Programming and critical analysis** (this code is the continuation of the code used in question 1)):

```
import matplotlib.pyplot as plt

#select the 2 features with the highest variance
Xnew = X.var().sort_values(ascending=False).head(2)

colors = [Y, y_pred]

#plot side by side
fig, ax = plt.subplots(1, 2, figsize=(10, 5))
for i in range(2):
    ax[i].scatter(X[Xnew.index[0]], X[Xnew.index[1]], c=colors[i])
    ax[i].set_xlabel(Xnew.index[0])
    ax[i].set_ylabel(Xnew.index[1])
    if i == 0:
        #background color
        ax[i].set_facecolor('lightblue')
        ax[i].set_title('Original Parkinson diagnosis')
    else:
        ax[i].set_title('Previously learned k=3 clusters (random = 0)')
plt.show()
```

Code used in question 4) of **II. Programming and critical analysis** (this code is the continuation of the code used in question 1) and 3)):

```
from sklearn.decomposition import PCA

#apply PCA
pca = PCA(n_components=0.8)
pca.fit(X)

#number of components
print("Number of components:", pca.n_components_)
```