TÉCNICO LISBOA

LEIC-T 2023/2024
Aprendizagem - Machine Learning
Homework 4
Deadline 30/10/2024 20:00
Submit on Fenix as pdf

I) (7 pts) Clustering

Given the data

$$\mathbf{x}_1 = \begin{pmatrix}2.5\\2.5\end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix}2\\2\end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix}0.5\\0.55\end{pmatrix},$$
$$\pi_1 = 0.6, \pi_2 = 0.4$$

$$C_1\left(u_1 = \begin{pmatrix}2\\2\end{pmatrix}, \Sigma_1 = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right),\ C_2\left(u_2 = \begin{pmatrix}0.5\\0.5\end{pmatrix}, \Sigma_2 = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right) .$$

i) (6 pts)

Perform one iteration of EM clustering algorithm step by step and determine the new parameters. Indicate all the calculations step by step. (To make the calculation easier for each step you can use a computer, however you should be able to do it by hand)

$x^{(1)}$:

• C = 1:

Likelihood: $\mathrm{p}\left(x^{(1)} \mid \mathrm{C} = 1\right) = \frac{1}{2\pi}\frac{1}{\det(\Sigma_1)}\ exp\left(-\frac{1}{2\pi}\left(x^{(1)} - u^{(1)}\right)^T\left(\Sigma_1\right)^{-1}\left(x^{(1)} - u^{(1)}\right)\right) =$

$= \frac{1}{2\pi}\frac{1}{1}\ exp(-0.0796) = 0.147$

Joint Probability: $\mathrm{p}\left(\mathrm{C} = 1,\ x^{(1)}\right) = \mathrm{p}(\mathrm{C} = 1)\ \mathrm{p}\left(x^{(1)} \mid \mathrm{C} = 1\right) = \pi_1 \times 0.147 = 0.0882$

• C = 2:

Likelihood: $\mathrm{p}\left(x^{(1)} \mid \mathrm{C} = 2\right) = \frac{1}{2\pi}\frac{1}{\det(\Sigma_2)}\ exp\left(-\frac{1}{2\pi}\left(x^{(1)} - u^{(2)}\right)^T\left(\Sigma_2\right)^{-1}\left(x^{(1)} - u^{(2)}\right)\right) =$

$= \frac{1}{2\pi}\frac{1}{1}\ exp(-1.2732) = 0.0446$

Joint Probability: $\mathrm{p}\left(\mathrm{C} = 2,\ x^{(1)}\right) = \mathrm{p}(\mathrm{C} = 2)\ \mathrm{p}\left(x^{(1)} \mid \mathrm{C} = 2\right) = \pi_2 \times 0.0446 = 0.0357$

• Posteriors:

C = 1: $\mathrm{p}\left(\mathrm{C} = 1 | x^{(1)}\right) = \frac{\mathrm{p}\left(\mathrm{C} = 1, x^{(1)}\right)}{\mathrm{p}\left(\mathrm{C} = 1, x^{(1)}\right) + \mathrm{p}\left(\mathrm{C} = 2, x^{(1)}\right)} = 0.7119$

C = 2: $\mathrm{p}\left(\mathrm{C} = 2 | x^{(1)}\right) = \frac{\mathrm{p}\left(\mathrm{C} = 2, x^{(1)}\right)}{\mathrm{p}\left(\mathrm{C} = 1, x^{(1)}\right) + \mathrm{p}\left(\mathrm{C} = 2, x^{(1)}\right)} = 0.2881$

$x^{(2)}$:

• C = 1:

Likelihood: $\mathrm{p}\left(x^{(2)} \mid \mathrm{C} = 1\right) = \frac{1}{2\pi}\frac{1}{\det(\Sigma_1)}\ exp\left(-\frac{1}{2\pi}\left(x^{(2)} - u^{(1)}\right)^T\left(\Sigma_1\right)^{-1}\left(x^{(2)} - u^{(1)}\right)\right) =$

$= \frac{1}{2\pi}\frac{1}{1}\ exp(0) = 0.1592$

Joint Probability: $\mathrm{p}\left(\mathrm{C} = 1,\ x^{(2)}\right) = \mathrm{p}(\mathrm{C} = 1)\ \mathrm{p}\left(x^{(2)} \mid \mathrm{C} = 1\right) = \pi_1 \times 0.1592 = 0.0955$

• C = 2:

Likelihood: $\mathrm{p}\left(x^{(2)} \mid \mathrm{C} = 2\right) = \frac{1}{2\pi}\frac{1}{\det(\Sigma_2)}\ exp\left(-\frac{1}{2\pi}\left(x^{(2)} - u^{(2)}\right)^T\left(\Sigma_2\right)^{-1}\left(x^{(2)} - u^{(2)}\right)\right) =$

$= \frac{1}{2\pi}\frac{1}{1}\ exp(-0.7162) = 0.0778$

Joint Probability: $\mathrm{p}\left(\mathrm{C} = 2,\ x^{(2)}\right) = \mathrm{p}(\mathrm{C} = 2)\ \mathrm{p}\left(x^{(2)} \mid \mathrm{C} = 2\right) = \pi_2 \times 0.\ 0778 = 0.0622$

• Posteriors:

C = 1: $\mathrm{p}\left(\mathrm{C} = 1 | x^{(2)}\right) = \frac{\mathrm{p}\left(\mathrm{C} = 1, x^{(2)}\right)}{\mathrm{p}\left(\mathrm{C} = 1, x^{(2)}\right) + \mathrm{p}\left(\mathrm{C} = 2, x^{(2)}\right)} = 0.6056$

C = 2: $\mathrm{p}\left(\mathrm{C} = 2 | x^{(2)}\right) = \frac{\mathrm{p}\left(\mathrm{C} = 2, x^{(2)}\right)}{\mathrm{p}\left(\mathrm{C} = 1, x^{(2)}\right) + \mathrm{p}\left(\mathrm{C} = 2, x^{(2)}\right)} = 0.3944$

$x^{(3)}$:

• C = 1:

Likelihood: $\mathrm{p}\left(x^{(3)} \mid \mathrm{C} = 1\right) = \frac{1}{2\pi}\frac{1}{\det(\Sigma_1)}\ exp\left(-\frac{1}{2\pi}\left(x^{(3)} - u^{(1)}\right)^T\left(\Sigma_1\right)^{-1}\left(x^{(3)} - u^{(1)}\right)\right) =$

$= \frac{1}{2\pi}\frac{1}{1}\ exp(-0.6927) = 0.0796$

Joint Probability: $\mathrm{p}\left(\mathrm{C} = 1,\ x^{(3)}\right) = \mathrm{p}(\mathrm{C} = 1)\ \mathrm{p}\left(x^{(3)} \mid \mathrm{C} = 1\right) = \pi_1 \times 0.1125 = 0.0478$

• C = 2:

Likelihood: $\mathrm{p}\left(x^{(3)} \mid \mathrm{C} = 2\right) = \frac{1}{2\pi}\frac{1}{\det(\Sigma_2)}\ exp\left(-\frac{1}{2\pi}\left(x^{(3)} - u^{(2)}\right)^T\left(\Sigma_2\right)^{-1}\left(x^{(3)} - u^{(2)}\right)\right) =$

$= \frac{1}{2\pi}\frac{1}{1}\ exp(-0.0004) = 0.1591$

Joint Probability: $\mathrm{p}\left(\mathrm{C} = 2,\ x^{(3)}\right) = \mathrm{p}(\mathrm{C} = 2)\ \mathrm{p}\left(x^{(3)} \mid \mathrm{C} = 2\right) = \pi_2 \times 0.1591 = 0.1273$

• Posteriors:

C = 1: $\mathrm{p}\left(\mathrm{C} = 1 | x^{(3)}\right) = \frac{\mathrm{p}\left(\mathrm{C} = 1, x^{(3)}\right)}{\mathrm{p}\left(\mathrm{C} = 1, x^{(3)}\right) + \mathrm{p}\left(\mathrm{C} = 2, x^{(3)}\right)} = 0.273$

C = 2: $\mathrm{p}\left(\mathrm{C} = 2 | x^{(3)}\right) = \frac{\mathrm{p}\left(\mathrm{C} = 2, x^{(3)}\right)}{\mathrm{p}\left(\mathrm{C} = 1, x^{(3)}\right) + \mathrm{p}\left(\mathrm{C} = 2, x^{(3)}\right)} = 0.727$

New parameters:

C = 1:

• likelihood:

$$u^1 = \frac{0.7119\binom{2.5}{2.5}+0.6056\binom{2}{2}+0.273\binom{0.5}{0.55}}{0.7119+0.6056+0.273} = \binom{1.9664}{1.9749}$$

$$\sum_{11}^{1} = \frac{0.7119(2.5-1.9664)(2.5-1.9664)+0.6056(2-1.9664)(2-1.9664)+0.273(0.5-1.9664)(0.5-1.9664)}{0.7119+0.6056+0.273}$$

= 0.497

$$\sum_{12}^{1} = \frac{0.7119(2.5-1.9664)(2.5-1.9749)+0.6056(2-1.9664)(2-1.9749)+0.273(0.5-1.9664)(0.55-1.9749)}{0.7119+0.6056+0.273}$$

= 0.4844

$\sum_{21}^{1} = \sum_{12}^{1} = 0.4844$

$$\sum_{22}^{1} = \frac{0.7119(2.5-1.9749)(2.5-1.9749)+0.6056(2-1.9749)(2-1.9749)+0.273(0.55-1.9749)(0.55-1.9749)}{0.7119+0.6056+0.273}$$

= 0.4722

$$\Sigma^1 = \begin{pmatrix} 0.497 & 0.4844 \\ 0.4844 & 0.4722 \end{pmatrix}$$

C = 2:

• likelihood:

$$u^2 = \frac{0.2881\binom{2.5}{2.5}+0.3944\binom{2}{2}+0.727\binom{0.5}{0.55}}{0.2881+0.3944+0.727} = \binom{1.4464}{1.4691}\binom{1.3286}{1.3543}$$

$$\sum_{11}^{2} = \frac{0.2881(2.5-1.3286)(2.5-1.3286)+0.3944(2-1.3286)(2-1.3286)+0.727(0.5-1.3286)(0.5-1.3286)}{0.2881+0.3944+0.727}$$

= 0.7607

$$\sum_{12}^{2} = \frac{0.2881(2.5-1.3286)(2.5-1.3543)+0.3944(2-1.3286)(2-1.3543)+0.727(0.5-1.3286)(0.55-1.3543)}{0.2881+0.3944+0.727}$$

= 0.7394

$\sum_{21}^{2} = \sum_{12}^{1} = 0.7394$

$$\sum_{22}^{2} = \frac{0.2881(2.5-1.3543)(2.5-1.3543)+0.3944(2-1.3543)(2-1.3543)+0.727(0.55-1.3543)(0.55-1.3543)}{0.2881+0.3944+0.727}$$

= 0.7186

$$\Sigma^2 = \begin{pmatrix} 0.7607 & 0.7394 \\ 0.7394 & 0.7186 \end{pmatrix}$$

New Priors:

$$p(C=1) = \frac{p(C=1|x^{(1)})+p(C=1|x^{(2)})+p(C=1|x^{(3)})}{p(C=1|x^{(1)})+p(C=1|x^{(2)})+p(C=1|x^{(3)})+p(C=2|x^{(1)})+p(C=2|x^{(2)})+p(C=2|x^{(3)})} = 0.5302$$

$$p(C=2) = \frac{p(C=1|x^{(1)})+p(C=1|x^{(2)})+p(C=1|x^{(3)})}{p(C=1|x^{(1)})+p(C=1|x^{(2)})+p(C=1|x^{(3)})+p(C=2|x^{(1)})+p(C=2|x^{(2)})+p(C=2|x^{(3)})} = 0.4698$$

ii) (1 pts)

Performing a hard assignment of observations to clusters identify the silhouette of the larger cluster

O maior cluster é o C_2. x_1 e x_2 estão atribuídos a C_2 pois os posterior values são menores quando $C = 2$.

$\mathbf{x}_1 = \begin{pmatrix} 2.5 \\ 2.5 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, x_3 = \begin{pmatrix} 0.5 \\ 0.55 \end{pmatrix}$

$$s(x_1) = 1 - \frac{\|x_1 - x_2\|_2}{\|x_1 - x_3\|_2} = 1 - \frac{0.7071}{2.7933} = 0.7469$$

$$s(x_2) = 1 - \frac{\|x_2 - x_1\|_2}{\|x_2 - x_3\|_2} = 1 - \frac{0.7071}{2.0863} = 0.6611$$

silhouette:

$$S(C_2) = \frac{0.7469 + 0.6611}{2} = 0.704$$

II Software Experiments (3pts)

a) (2 pts)
Download the jupyter notebook HM4_CL.ipynb. Load the build in data set “wine” preform kmeans and EM clustering with 2, 3, 4 cluster and indicate the silhouette as defined in the notebook for each experiment. Which k-value give the ideal value.

k-means:
2 cluster - silhouette = 0.655521358978658
3 cluster - silhouette = 0.5711381937868838
4 cluster - silhouette = 0.5587089480903824

EM:
2 cluster - silhouette = 0.5643242782521394
3 cluster - silhouette = 0.34726590057721557
4 cluster - silhouette = 0.32884029199766257

A maior silhouette é obtida sempre que temos 2 clusters, assim sendo podemos dizer que o k-value ideal é 2.

Perform PCA with two components with 2, 3, 4 cluster and indicate the silhouette as defined in the notebook for each experiment. Which k-value give the ideal value? Is the ideal k value the same with PCA and without?

PCA:

k-means:
2 cluster - silhouette = 0.6572176888364498
3 cluster - silhouette = 0.5716547257508234
4 cluster - silhouette = 0.5589476208089745

EM:
2 cluster - silhouette = 0.6507063215091908
3 cluster - silhouette = 0.28006565487155105
4 cluster - silhouette = 0.10854149525298863

A maior silhouette é obtida sempre que temos 2 clusters, assim sendo podemos dizer que o k-value ideal é 2.

Assim, podemos concluir que o k-value ideal (2) é o mesmo com e sem PCA

b) (1 pts)

Load the build in data set “breast_cancer” perform k-means and EM clustering with 2 cluster and indicate the silhouette as defined in the notebook. Which one is better?

Without PCA:
k-means:
silhouette = 0.6972646156059464
EM:
Silhouette = 0.5315172918032405

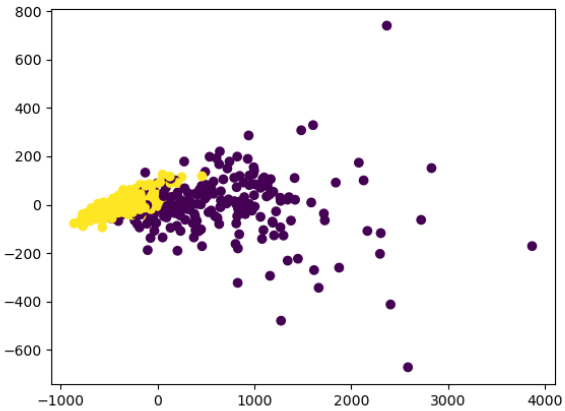
With PCA:
k-means:
silhouette = 0.6984195775999954
EM:
Silhouette = 0.5865823748565954

Com base nestes valores é possível perceber que em ambos os casos, o k-means tem uma silhueta mais alta em comparação com o EM clustering. Portanto, com base no valor da silhueta, o k-means parece ter um melhor desempenho no conjunto de dados "breast_cancer", com ou sem PCA. K-means fornece clusters mais bem separados e coesos.

Perform PCA with two components with 2, cluster and indicate the silhouette as defined in the notebook. Plot the scatter plot. When you compare the plots and the silhouette values and look at the scatter plot of the PCA mapped data, what is your conclusion. Short, one sentence pls.

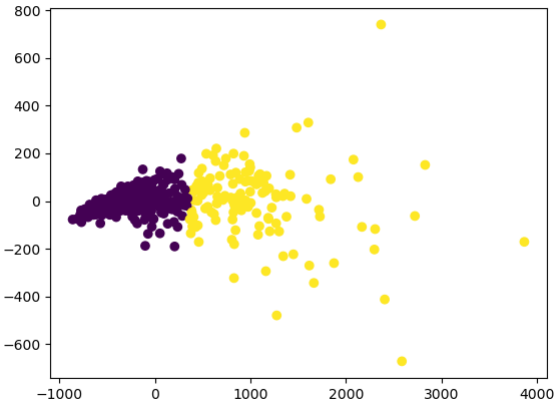
Com PCA

Scatter:



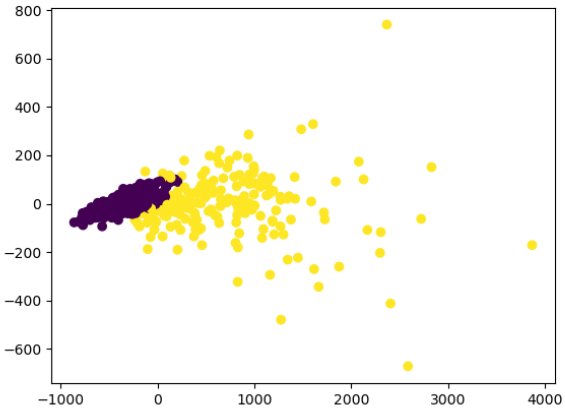
k-means:

silhouette = 0.6984195775999954



EM:

Silhouette = 0.5865823748565954



É possível concluir através destes plots que, ainda que a silhueta do k-means pareça ter um melhor desempenho neste conjunto de dados, o plot com o EM clustering é uma representação mais fiel do scatter plot com PCA.