DDA3020 Machine Learning Spring 2023
Produced by TA Teams

Answers for Assignment 4

1 Problem 1 EM

Proof

The distribution and the log-likelihood of a mixture of Bernoullis are given by:

$$p(x_i|\mu_k) = \prod_{j=1}^{D} \mu_{kj}^{x_{ij}} (1-\mu_{kj})^{1-x_{ij}}$$
$$\log p(x_i|\mu_k) = \sum_{j=1}^{D} x_{ij} \log \mu_{kj} + (1-x_{ij}) \log(1-\mu_{kj})$$

Therefore, we get the following auxiliary function:

$$\begin{aligned}
Q(\mu, \mu^{t-1}) &= E\left[\sum_{i=1}^{N} \log p(x_i, z_i|\mu^{t-1})\right] \\
&= \sum_i \sum_k r_{ik} \log \pi_k + \sum_i \sum_k r_{ik} \log p(x_i|\mu_k) \\
&= \sum_i \sum_k r_{ik} \log \pi_k + \sum_i \sum_k r_{ik} \sum_j (x_{ij} \log \mu_{kj} + (1-x_{ij}) \log(1-\mu_{kj}))
\end{aligned}$$

In M step, μ_{kj} is given by:

$$\mu_{kj} = \arg\max_{\mu_{kj}} Q(\mu, \mu^{t-1})$$

Let's take the derivative of $Q(\mu, \mu^{t-1})$ w.r.t. μ_{kj}.

$$\frac{\partial Q}{\partial \mu_{kj}} = \sum_i r_{ik} \left(\frac{x_{ij}}{\mu_{kj}} - \frac{1-x_{ij}}{1-\mu_{jk}}\right) \tag{1}$$
$$= \frac{1}{\mu_{kj}(1-\mu_{kj})} \left(\sum_i r_{ik} x_{ij} - \mu_{kj} \sum_i r_{ik}\right) \tag{2}$$

Then, by derivative goes to 0.

$$\mu_{jk} = \frac{\sum_i r_{ik} x_{ij}}{\sum_i r_{ik}}$$

Beta prior is given by:

$$\beta(\alpha, \beta, \mu_{kj}) = A\mu_{kj}^{\alpha-1}(1-\mu_{kj})^{\beta-1} \quad \left(\because A = \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!}\right)$$
$$\log \beta(\alpha, \beta, \mu_{kj}) = \log \beta(\alpha, \beta, \mu_{kj})$$
$$= \log A + (\alpha-1)\log \mu_{kj} + (\beta-1)\log(1-\mu_{kj})$$

In M step of MAP estimation, μ_{kj} is given by:

$$\mu_{kj} = \arg\max_{\mu_{kj}} Q' := Q(\mu, \mu^{t-1}) + \sum_j \sum_k \log \beta(\alpha, \beta, \mu_{kj}) \tag{3}$$

The derivative of $\log \beta(\alpha, \beta, \mu_{kj})$ w.r.t. μ_{kj} is:

$$\frac{\partial}{\partial \mu_{kj}} \log \beta(\alpha, \beta, \mu_{kj}) = \frac{\alpha-1}{\mu_{kj}} - \frac{\beta-1}{1-\mu_{kj}} \tag{4}$$
$$= \frac{\alpha - 1 - \mu_{kj}(\alpha+\beta-2)}{\mu_{kj}(1-\mu_{kj})} \tag{5}$$

By combining equations 2,3,5 and $\frac{\partial Q'}{\partial \mu_{kj}} = 0$, we get the following equation:

$$\mu_{kj} = \frac{(\sum_i r_{ik}x_{ij}) + \alpha - 1}{(\sum_i r_{ik}) + \alpha + \beta - 2}$$

2 Problem 2 Kmeans

1. - The 1st iteration:
     epoch 1− start:
     A1:
     d(A1, seed 1) = 0 as A1 is seed 1
     d(A1, seed 2) = $\sqrt{13} > 0$
     d(A1, see 3) = $\sqrt{65} > 0 \Rightarrow$ A1 $\in$ cluster 1
     A2:

     d(A2, seed 1) = $\sqrt{25} = 5$
     d(A2, seed 2) = $\sqrt{18} = 4.24$
     d(A2, seed 3) = $\sqrt{10} = 3.16$ $\leftarrow$ smaller
     $\rightarrow$ A2 $\in$ cluster 3

     A3:

     d(A3, seed 1) = $\sqrt{36} = 6$
     d(A3, seed 2) = $\sqrt{25} = 5$ $\leftarrow$ smaller
     d(A3, seed 3) = $\sqrt{53} = 7.28$
     $\rightarrow$ A3 $\in$ cluster 2

A4:

$$\text{d(A4, seed 1)} = \sqrt{13}$$
$$\text{d(A4, seed 2)} = 0 \text{ as } A4 \text{ is seed 2}$$
$$\text{d(A4, seed 3)} = \sqrt{52} > 0$$
$$\Rightarrow \text{A4} \in \text{ cluster 2}$$

A5:

$$\text{d}(\text{A5}, seed1) = \sqrt{50} = 7.07$$
$$\text{d(A5, seed 2)} = \sqrt{13} = 3.60 \leftarrow \text{ smaller}$$
$$\text{d(A5, seed 3)} = \sqrt{45} = 6.70$$
$$\Rightarrow \text{A5} \in \text{ cluster 2}$$

A6:

$$(A6, seed1) = \sqrt{52} = 7.21$$
$$\text{d(A6, seed 2)} = \sqrt{17} = 4.12 \leftarrow \text{ smaller}$$
$$\text{d(A6, seed 3)} = \sqrt{29} = 5.38$$
$$\rightarrow \text{A6} \in \text{ cluster2}$$

A7:

$$\text{d(A7, seed 1)} = \sqrt{65} > 0$$
$$\text{d(A7, seed 2)} = \sqrt{52} > 0$$
$$\text{d(A7, seed 3)} = 0 \text{ as A7 is seed 3}$$
$$\Rightarrow \text{A7} \in \text{ cluster3}$$

A8:

$$\text{d(A8, seed 1)} = \sqrt{5}$$
$$\text{d(A8, seed 2)} = \sqrt{2} \text{ \& smaller}$$
$$\text{d(A8, seed 3)} = \sqrt{58}$$
$$\Rightarrow \text{A8} \in \text{ cluster 2}$$

end of epoch1

new clusters: 1: $\{\text{A1}\}, 2 : \{\text{A3, A4, A5, A6, A8}\}, 3 : \{\text{A2, A7}\}$

centers of the new clusters: C1 $= (2, 10)$

C2 $= ((8 + 5 + 7 + 6 + 4)/5, (4 + 8 + 5 + 4 + 9)/5) = (6, 6)$

C3 $= ((2 + 1)/2, (5 + 2)/2) = (1.5, 3.5)$

We would need two more epochs to get the convergent result.
After the 2^{nd} epoch the results would be: 1: $\{A1, A8\}, 2 : \{A3, A4, A5, A6\}, 3 : \{A2, A7\}$ with centers C1 $= (3, 9.5)$, C2 $= (6.5, 5.25)$ and C3 $= (1.5, 3.5)$. After the

3^{rd} epoch, the results would be: 1: $\{A1, A4, A8\}, 2 : \{A3, A5, A6\}, 3 : \{A2, A7\}$ with centers $C1 = (3.66, 9), C2 = (7, 4.33)$ and $C3 = (1.5, 3.5)$. The final clusters are shown below.

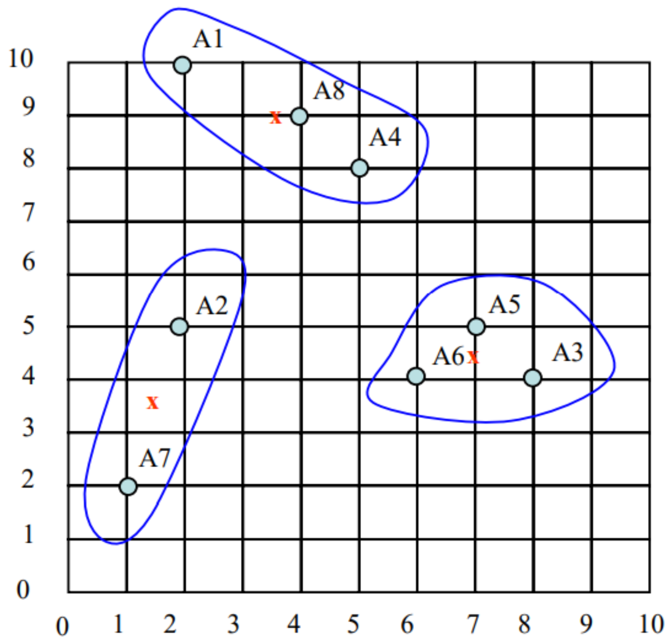

Figure 1:

2. Consider the constriants: epoch 1− start:
   A1:
   d(A1, seed 1) = 0 as A1 is seed 1
   d(A1, seed 2) = $\sqrt{13} > 0$
   d(A1, see 3) = $\sqrt{65} > 0 \Rightarrow$ A1 ∈ cluster 1
   A2:

$$\text{d(A2, seed 1)} = \sqrt{25} = 5$$
$$\text{d(A2, seed 2)} = \sqrt{18} = 4.24$$
$$\text{d(A2, seed 3)} = \sqrt{10} = 3.16 \quad \leftarrow \text{ smaller}$$
$$\rightarrow \text{A2} \in \text{ cluster 3}$$

   A3:

$$\text{d(A3, seed 1)} = \sqrt{36} = 6$$
$$\text{d(A3, seed 2)} = \sqrt{25} = 5 \quad \leftarrow \text{ smaller}$$
$$\text{d(A3, seed 3)} = \sqrt{53} = 7.28$$
$$\rightarrow \text{A3} \in \text{ cluster 2}$$

A4:

$$\begin{aligned}&\mathrm{d}(\mathrm{A4},\ \text{seed } 1)=\sqrt{13}\\&\mathrm{d}(\ \mathrm{A4},\ \text{seed } 2)=0 \text{ as } A4 \text{ is seed } 2\\&\mathrm{d}(\ \mathrm{A4},\ \text{seed } 3)=\sqrt{52}>0\\&\Rightarrow \mathrm{A4}\in \text{ cluster } 2\end{aligned}$$

A5:

$$\begin{aligned}&\mathrm{d}(\mathrm{A5}, seed1)=\sqrt{50}=7.07\\&\mathrm{d}(\mathrm{A5},\ \text{seed } 2)=\sqrt{13}=3.60\leftarrow \text{ smaller}\\&\mathrm{d}(\mathrm{A5},\ \text{seed } 3)=\sqrt{45}=6.70\\&\Rightarrow \mathrm{A5}\in \text{ cluster } 2\end{aligned}$$

However, $A5$ has a cannot link with $A4$, $\Rightarrow$ A5 $\in$ cluster 3
A6:

$$\begin{aligned}&(A6, seed1)=\sqrt{52}=7.21\\&\mathrm{d}(\mathrm{A6},\ \text{seed } 2)=\sqrt{17}=4.12\leftarrow \text{ smaller}\\&\mathrm{d}(\mathrm{A6},\ \text{seed } 3)=\sqrt{29}=5.38\\&\rightarrow \mathrm{A6}\in \text{ cluster2}\end{aligned}$$

However, A6 has a must link with A2, $\Rightarrow$ A6 $\in$ cluster 3.
A7:

$$\begin{aligned}&\mathrm{d}(\mathrm{A7},\ \text{seed } 1)=\sqrt{65}>0\\&\mathrm{d}(\ \mathrm{A7},\ \text{seed } 2)=\sqrt{52}>0\\&\mathrm{d}(\ \mathrm{A7},\ \text{seed } 3)=0 \text{ as A7 is seed } 3\\&\Rightarrow \mathrm{A7}\in \text{ cluster3}\end{aligned}$$

A8:

$$\begin{aligned}&\mathrm{d}(\mathrm{A8},\ \text{seed } 1)=\sqrt{5}\\&\mathrm{d}(\ \mathrm{A8},\ \text{seed } 2)=\sqrt{2}\leftarrow \text{smaller}\\&\mathrm{d}(\mathrm{A8},\ \text{seed } 3)=\sqrt{58}\\&\Rightarrow \mathrm{A8}\in \text{ cluster } 2\end{aligned}$$

However, A8 has a must link with A1, $\Rightarrow$ A8 $\in$ cluster 1. end of epoch1
new clusters: 1: $\{\mathrm{A1}, \mathrm{A8}\}, 2:\{\mathrm{A3},\ \mathrm{A4}\}, 3:\{\mathrm{A2},\ \mathrm{A7}\ \ \mathrm{A5},\ \mathrm{A6}\}$

centers of the new clusters: C1 $=(3, 9.5)$
C2 $=(6.5, 6)$
C3 $=(4, 4)$

By the iterations, we can get many possible clusters.

3 Problem 3 PCA

Solution: Here is an example, which is not unique.

The mean of the data is [6.9, 3.5, 5.1]. The covariance matrix of centered $\overline{X}$ is given by

$$\begin{bmatrix} 2.32 & 1.61 & -0.43 \\ 1.61 & 2.50 & -1.28 \\ -0.43 & -1.28 & 7.88 \end{bmatrix}$$

Then the eigenvalue and their corresponding eigenvectors are: $\lambda_1 = 8.2761$ with [0.1369, 0.2505, -0.9584] $\lambda_2 = 3.6747$ with [0.6991, 0.6610, -0.2727] $\lambda_3 = 0.7493$ with [0.7018, 0.7073, -0.0846]

Choose the last two largest eigenvalues for PCA and their corresponding eigenvectors to form U, we have

$$Z = U^T\overline{X} = \begin{bmatrix} 2.15 & -3.80 & -0.15 & 4.71 & -1.29 & -4.10 & 1.63 & -2.12 & 0.23 & 2.75 \\ -0.17 & -2.89 & -0.99 & 1.30 & 2.28 & 0.14 & -2.23 & 3.25 & 0.37 & -1.07 \end{bmatrix}$$

where each column is the new projection.