CSE/ISYE 6740 Mid-term Exam 2 (Fall 2019)

Le Song

10/24 Thr, 12:00 - 1:15 pm

- Name:
- GT Username:
- E-mail:

Problem	Point	Your Score
1	25	
2	25	
3	25	
4	25	
Total	100	

Instructions:

- Try your best to be clear as much as possible. No credit may be given to unreadable writing.
- The exam is open note, but **no electronic devices** (including smart phones) are allowed.

1 Bayes' Rule and Maximum Likelihood [25 pts]

(a) There is a coin but you don't know whether it is a normal coin (hypothesis $\boldsymbol{A}$, $P(\text{Head}|\boldsymbol{A}) = P(\text{Tail}|\boldsymbol{A}) = 0.5$) or a double-headed coin (hypothesis $\boldsymbol{B}$, $P(\text{Head}|\boldsymbol{B}) = 1$). If you toss a coin, a norm one has 50% chance to land heads up. Assuming your initial belief is,

$$p(\boldsymbol{A}) = 0.8, \quad p(\boldsymbol{B}) = 0.2.$$

You toss the coin twice, and observe that the coin always shows its head when it lands.

(i) Find the likelihoods of the tossing result given hypothesis $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. [6 pts]

Solution: The likelihoods are,

$$\begin{aligned} P(\text{First} = \text{Head}, \text{Second} = \text{Head}|\boldsymbol{A}) &= P(\text{First} = \text{Head}|\boldsymbol{A}) \cdot P(\text{Second} = \text{Head}|\boldsymbol{A}) \\ &= P(\text{Head}|\boldsymbol{A})^2 \\ &= 0.25, \end{aligned}$$

$$\begin{aligned} P(\text{First} = \text{Head}, \text{Second} = \text{Head}|\boldsymbol{B}) &= P(\text{First} = \text{Head}|\boldsymbol{B}) \cdot P(\text{Second} = \text{Head}|\boldsymbol{B}) \\ &= P(\text{Head}|\boldsymbol{B})^2 \\ &= 1. \end{aligned}$$

(ii) What are your updated beliefs for $\boldsymbol{A}$ and $\boldsymbol{B}$ after observing the result? [7 pts]

Solution: Using Bayes' theorem, we have the updated beliefs,

$$\begin{aligned} P(\boldsymbol{A}|\text{Head}^2) &= \frac{P(\text{Head}^2|\boldsymbol{A}) \cdot P(\boldsymbol{A})}{P(\text{Head}^2|\boldsymbol{A}) \cdot P(\boldsymbol{A}) + P(\text{Head}^2|\boldsymbol{B}) \cdot P(\boldsymbol{B})} \\ &= \frac{0.25 \times 0.8}{0.25 \times 0.8 + 1 \times 0.2} \\ &= 0.5, \end{aligned}$$

$$P(\boldsymbol{B}|\text{Head}^2) = 1 - P(\boldsymbol{A}|\text{Head}^2) = 0.5,$$

where we use Head^2 as a short-hand notation for $\text{First} = \text{Head}, \text{Second} = \text{Head}$.

(b) The independent random variables $X_1, X_2, ..., X_n$ follow the Poisson distribution

$$P(X_i = k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \cdots,$$

where the parameter λ is positive. Find the MLE of λ. [12 pts]

Solution: We first find the log-likelihood,

$$\begin{aligned}
\ell(\lambda) &= \log \prod_{i=1}^{n} P(X_i = k_i|\lambda) \\
&= \sum_{i=1}^{n} \log \frac{\lambda^{k_i} e^{-\lambda}}{k_i!} \\
&= \sum_{i=1}^{n} \big(k_i \log \lambda - \lambda - \log(k_i!)\big) \\
&= \sum_{i=1}^{n} k_i \log \lambda - n\lambda - \sum_{i=1}^{n} \log(k_i!).
\end{aligned}$$

Then we take its derivative w.r.t. λ,

$$\frac{\partial \ell}{\partial \lambda} = \frac{\sum_{i=1}^{n} k_i}{\lambda} - n.$$

By setting it to 0, we have,

$$\widehat{\lambda}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} k_i.$$

2 Clustering [25 pts]

(a) Consider five points in the space: x_1 =(1,2), x_2 =(0,0), x_3 =(-1, 2), x_4 =(3,0) and x_5 =(-2,2). You want to apply the K-means algorithm with K=2 and with **Euclidean** distance. Assuming that you initialize the cluster as $u_1 = x_2$ and $u_2 = x_5$. Compute the coordinates of the centroids after one iteration of K-means. [15 pt]

Solution: first we compute the distance for cluster assignment

$$d_{21} = \sqrt{5}, d_{23} = \sqrt{5}, d_{24} = 3$$
$$d_{51} = 3, d_{53} = 1, d_{54} = \sqrt{29},$$

thus

$$c_1 = \{x_1, x_2, x_4\}$$
$$c_2 = \{x_3, x_5\}.$$

Compute the centroids with the above assignments

$$u_1 = \frac{1}{3}(x_1 + x_2 + x_4) = (\frac{4}{3}, \frac{2}{3}) = (1.33, 0.66)$$
$$u_2 = \frac{1}{2}(x_3 + x_5) = (\frac{-3}{2}, 2) = (-1.5, 2).$$

(b) Compute the loss function $J = \frac{1}{n}\sum_i^n \|x_i - u_{c_i}\|$ before and after the iteration, where c_i denotes the cluster that x_i belongs to. [10 pts]

before the iteration we have assignment

$$\begin{aligned} c_1 &= \{x_1, x_2, x_4\} \\ c_2 &= \{x_3, x_5\}. \end{aligned}$$

So the loss is

$$\begin{aligned} J &= \frac{1}{5}(d_{21} + d_{24} + d_{53}) \\ &= \frac{1}{5}(\sqrt{5} + 3 + 1) \\ &= \frac{\sqrt{5} + 4}{5} \\ &= 1.247(or\ \frac{15}{5} = 3\ if\ using \| \cdot \|^2) \end{aligned}$$

After the iteration, we, again, compute the distance

$$\begin{aligned} d_{u_1 1} &= \frac{\sqrt{17}}{3}, d_{u_1 2} = \frac{2\sqrt{5}}{3}, d_{u_1 3} = \frac{\sqrt{65}}{3}, d_{u_1 4} = \frac{\sqrt{29}}{3}, d_{u_1 5} = \frac{\sqrt{116}}{3} \\ d_{u_2 1} &= \frac{5}{2}, d_{u_2 2} = \frac{5}{2}, d_{u_2 3} = \frac{1}{2}, d_{u_2 4} = \frac{\sqrt{97}}{2}, d_{u_2 5} = \frac{1}{2}. \end{aligned}$$

Then the assignment becomes

$$\begin{aligned} c_1 &= \{x_1, x_2, x_4\} \\ c_2 &= \{x_3, x_5\}. \end{aligned}$$

So the loss is

$$\begin{aligned} J &= \frac{1}{5}(d_{u_1 1} + d_{u_1 2} + d_{u_1 4} + d_{u_2 3} + d_{u_2 5}) \\ &= \frac{1}{5}(\frac{\sqrt{17}}{3} + \frac{2\sqrt{5}}{3} + \frac{\sqrt{29}}{3} + \frac{1}{2} + \frac{1}{2}) \\ &= \frac{2\sqrt{17} + 4\sqrt{5} + 2\sqrt{29} + 6}{30} \\ &= 1.132(or\ 1.566\ if\ using \| \cdot \|^2) \end{aligned}$$

3 Principal Component Analysis [25 pts]

We are given a matrix

$$X = \begin{bmatrix} 6 & -4 \\ -3 & 5 \\ -2 & 6 \\ 7 & -3 \end{bmatrix}$$

where each row is a data point. We need to use PCA to reduce the dimension of X from 2 to 1.

(a) As a first step, compute the covariance matrix for the sample points? (Note: X is not centered.) [8 pts]

Solution: The centered matrix is

$$X_c = \begin{bmatrix} 4 & -5 \\ -5 & 4 \\ -4 & 5 \\ 5 & -4 \end{bmatrix}$$

and the corresponding covariance matrix is

$$X_c^T X_c = \begin{bmatrix} 82 & -80 \\ -80 & 82 \end{bmatrix}$$

(b) Compute the **unit** eigenvectors and corresponding eigenvalues of the obtained covariance matrix. [9 pts] Solution: The unit eigenvectors are

$$e_0 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

with eigenvalue 2 and

$$e_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

with eigenvalue 162.

(c) We now use PCA to project the sample points onto a one-dimensional space. What one-dimensional subspace are we projecting onto? For each of the 4 points in X write the coordinate (in principal coordinate space) that the point is projected to? [8 pts]

Solution: We are projecting onto the subspace spanned by e_1. For instance, the projection of point $x_1 = (6, -4)$ will be $< x_1, e_1 >= \frac{6}{\sqrt{2}} + \frac{4}{\sqrt{2}}$. The corresponding projections are

$$projections = \begin{bmatrix} 10/\sqrt{2} \\ -8/\sqrt{2} \\ -8/\sqrt{2} \\ 10/\sqrt{2} \end{bmatrix}$$

4 EM and Information [25 pts]

Suppose we have a bunch of coins $\mathcal{C}$ consisting two kinds of coins. Mathematically, it obeys a mixed Bernoulli distribution:

$$X \sim F = \pi F_1 + (1-\pi)F_2$$

where $\pi \in [0,1]$ and $F_1 = \mathrm{Ber}(p_1)$, $F_2 = \mathrm{Ber}(p_2)$. That's to say, the kind of the coin is determined by a latent variable $Z = 1$ or $Z = 2$. Here $Ber(p)$ means the coin gives 1 (head) with probability p and gives 0 (tail) with probability $1-p$.

We initialize parameters $p_1 = \frac{1}{3}$, $p_2 = \frac{2}{3}$ and $\pi = \frac{1}{2}$. Now, we draw 4 coins X_1, X_2, X_3, X_4 independently from $\mathcal{C}$ and have 3 independent trials for each of them. The result shows:

Coins	X_1	X_2	X_3	X_4
Trail 1	1	0	0	1
Trail 2	1	1	1	0
Trail 3	0	1	0	1

(a) Use EM algorithm for **one** step, we update $F = F(p_1 = \frac{1}{3}, p_2 = \frac{2}{3}, \pi = \frac{1}{2})$ to $F'(p_1', p_2', \pi')$. Write down your EM algorithm and show the value of p_1', p_2', π'. [15 pts].

Answer: In E step, let $\tau_j^{(i)} = \mathbb{P}(X_i \sim F_j)$, we have:

$$\tau_1^{(1)} = \frac{\pi p_1^2(1-p_1)}{\pi p_1^2(1-p_2) + (1-\pi)p_2^2(1-p_2)} = \frac{1}{3}$$

$$\tau_1^{(3)} = \frac{\pi p_1(1-p_1)^2}{\pi p_1^2(1-p_2) + (1-\pi)p_2(1-p_2)^2} = \frac{2}{3}$$

and $\tau_1^{(2)} = \tau_1^{(4)} = \tau_1^{(1)} = \frac{1}{3}$, $\tau_2^{(i)} = 1 - \tau_1^{(i)}$

In M step, denote all parameters as Θ, let $\pi_1 = \pi, \pi_2 = 1-\pi$, X_i is the summation of three trails. Then, we want to optimize $Q(\Theta, \Theta_{\text{old}})$:

$$\begin{aligned} Q(\Theta, \Theta_{\text{old}}) &= \sum_{i=1}^{4}\sum_{j=1}^{2} \tau_j^{(i)} \log\left(\pi_j p_j^{X_i}(1-p_j)^{3-X_i}\right) \\ &= 3[\frac{1}{3}\log\left(3\pi p_1^2(1-p_1)\right) + \frac{2}{3}\log\left(3(1-\pi)p_2^2(1-p_2)\right)] \\ &\quad + [\frac{2}{3}\log\left(3\pi p_1(1-p_1)^2\right) + \frac{1}{3}\log\left(3(1-\pi)p_2(1-p_2)^2\right)] \end{aligned}$$

Let the gradient equals to zero, we have:

$$\frac{5}{3\pi'} = \frac{7}{3(1-\pi')}$$

$$\frac{8}{3p_1'} = \frac{7}{3(1-p_1')}$$

$$\frac{13}{3p_2'} = \frac{8}{3(1-p_2')}$$

Hence, $\pi' = \frac{5}{12}, p_1' = \frac{8}{15}, p_2' = \frac{13}{21}$

show algorithm: 5 pts, E-step: 5pts, M-step: 5pts

(b) Using $F'(\pi', p_1', p_2')$, what is the Entropy of X [10 pts].

Remark: If you don't know $F'(p_1', p_2', \pi')$, you can use $F(p_1, p_2, \pi)$ and get half credits.

Answer: Using the result in (a), we have the distribution of X is $\mathrm{Ber}(p)$, where

$$p = \pi p_1 + (1-\pi)p_2 = \frac{2}{9} + \frac{13}{36} = \frac{7}{12}$$

So, the entropy is:

$$H_{F'}(X) = -[p \log p + (1-p)\log(1-p)] = -\frac{5}{12}\log\frac{5}{12} - \frac{7}{12}\log\frac{7}{12}$$

If you use distribution F, X is a fair coin. Hence, $H_F(X) = \log 2$