STATS303 Midterm Exam

Spring'22 Session 4

NAME: ______________________________

Perferred Name / Name in Chinese (optional): ______________________________

NetID: ______________________________

Time: 75min

Total points: 300pt

Note:

- Please keep this exam confidential and do not share the problems with others.
- This exam is open-book and open-notes. However, online search or online help is strictly banned.

Problem 1. (90pt)

1. Suppose we want to estimate the mean μ and the variance σ^2 of a Gaussian density $\mathcal{N}(\mu, \sigma^2)$.

   (a) (20pt) Is the maximum-likelihood estimator for μ biased or unbiased? You don't need to provide your reason.

   (b) (20pt) Is the maximum-likelihood estimator for σ^2 biased or unbiased? You don't need to provide your reason.

2. (20pt) Suppose f and g are two different estimators for the same parameter $\boldsymbol{\theta}$ in some density $p(\boldsymbol{x}|\boldsymbol{\theta})$. Is it possible that the bias of f is smaller than the bias of g, while at the same time the variance of f is also smaller than the variance of g? If yes, give an example of such f and g; otherwise, briefly state the reason.

3. Let X be a random variable. $X \sim \text{Unif}(\theta)$ with the density

$$p(x) = \begin{cases} \dfrac{1}{\theta}, & \text{if } 0 \leq x \leq \theta \\ \\ 0, & \text{otherwise .} \end{cases}$$

   (a) (20pt) Consider the sample $\mathcal{X} = \{x_n\}_{n=1}^{N}$ where $x_n \sim^{i.i.d.} \text{Unif}(\theta)$. For the parameter θ above, write the likelihood $l(\theta|\mathcal{X})$ and the log-likelihood $\mathcal{L}(\theta|\mathcal{X})$.

   (b) (10pt) Find the maximum likelihood estimator $\hat{\theta}_{\text{MLE}}$.

Problem 2. (110pt)

1. (30pt) Consider applying K-means with $K = 2$ clusters to the four data points

$$\boldsymbol{x}_1 = (0,0)^{\mathrm{T}}, \boldsymbol{x}_2 = (2,0)^{\mathrm{T}}, \boldsymbol{x}_3 = (0,3)^{\mathrm{T}}, \boldsymbol{x}_4 = (2,3)^{\mathrm{T}}$$

in $\mathbb{R}^2$. Suppose the initial centers are set to be $\boldsymbol{\mu}_1 = (0,0)^{\mathrm{T}}$ and $\boldsymbol{\mu}_2 = (4,3)^{\mathrm{T}}$. Write the E-step and the M-step for the first iteration. You need to clearly state the locations of the centers and the labels of the data points.

2. Consider same data points as in Part 1. Denote $\boldsymbol{X} = \{\boldsymbol{x}_n\}_{n=1}^{4}$. Suppose we fit a Gaussian Mixture Model (GMM) to the data set using the *general* EM algorithm. Answer the following questions.

   (a) (20pt) According to the general EM algorithm, is $\boldsymbol{X}$ the complete dataset or not? If not, what would make it complete?

   (b) (20pt) In the context of the general EM algorithm, briefly explain the relationship between "expectation" and "responsibility".

   (c) (40pt) Suppose in the current iteration, the two Gaussians are given by $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{I})$ and $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{I})$, where $\boldsymbol{\mu}_1 = (0, 0)^{\mathrm{T}}$ and $\boldsymbol{\mu}_2 = (4, 3)^{\mathrm{T}}$. Moreover, $\pi_1 = \pi_2 = 0.5$. What are the values of the "responsibilities" for $\boldsymbol{x}_1$, the first data point? Your answer may contain exponential functions.

Problem 3. (80pt)

Suppose we are given a dataset $\mathcal{X} = \{\boldsymbol{x}_n, y_n\}_{n=1}^4$ where the inputs (independent variables) are

$$\boldsymbol{x}_1 = (0,0)^{\mathrm{T}}, \boldsymbol{x}_2 = (2,0)^{\mathrm{T}}, \boldsymbol{x}_3 = (0,2)^{\mathrm{T}}, \boldsymbol{x}_4 = (1,1)^{\mathrm{T}}$$

and the target values (dependent variables) are

$$y_1 = y_2 = y_3 = 0,\ y_4 = 1.$$

We want to perform a kernel ridge regression to this dataset.

1. (20pt) Explain briefly why ridge regression can be regarded as a Bayesian approach.
2. (60pt) Suppose we choose the polynomial kernel K defined by

$$K(\boldsymbol{x}, \boldsymbol{y}) = \left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} + 1\right)^2.$$

   Perform the kernel ridge regression to predict the target value for the input $\boldsymbol{x} = (2,2)^{\mathrm{T}}$ with the regularization parameter $\lambda = 1$. Your answer may contain vector and matrix operations and you don't need to do the computation.

Problem 4. (20pt)

Recall, that given a log-likelihood function $\mathcal{L}(\boldsymbol{\theta}|\mathcal{X}) = \sum_{n=1}^{N} \log p(\boldsymbol{x}_n|\boldsymbol{\theta})$, the score function is defined to be

$$\mathcal{S}(\boldsymbol{\theta}|\mathcal{X}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta}|\mathcal{X})}{\partial \boldsymbol{\theta}}.$$

Prove: $\mathbb{E}\left[\mathcal{S}(\boldsymbol{\theta}|\mathcal{X})\right] = 0.$

Bonus survey (5pt for completion):

On a scale of 1 to 9 (1 means easiest and 9 means the most difficult), *assuming the average difficulty of the homework problems is 5*, rate the difficulty of this exam. Circle a number below.

1 2 3 4 5 6 7 8 9

–END OF EXAM–