CS 771A: Intro to Machine Learning, IIT Kanpur			**Endsem Exam** (22 Nov 2022)
Name			**50 marks**
Roll No		Dept.	Page **1** of **6**

Instructions:
1. This question paper contains 3 pages (6 sides of paper). Please verify.
2. Write your name, roll number, department in **block letters** with **ink** on **each page**.
3. Write your final answers neatly **with a blue/black pen**. Pencil marks may get smudged.
4. Don't overwrite/scratch answers especially in MCQ.

Q1. Write **T** or **F** for True/False **in the box** and give justification below. **(4 x (1+2) = 12 marks)**

1	The Nikola company shares have a 40% chance of crashing if its owner Ksümnöle tweets something silly. The shares have a 10% chance of crashing if no silly tweet is sent. Ksümnöle tweets something silly with a 20% chance. Then, the probability that Nikola shares will crash, is less than 20%. Justify by calculating the probability.	
2	Given three vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^2$ such that $\mathbf{x}^\top \mathbf{z} > \mathbf{x}^\top \mathbf{y}$, it is always the case that $\|\mathbf{x} - \mathbf{z}\|_2^2 < \|\mathbf{x} - \mathbf{y}\|_2^2$. Give a proof if True else give a counter example.	
3	Consider the set $\mathcal{X} = \{-1, +1\}^3$ of 3D vectors with ± 1 coordinates. Any map $\phi: \mathcal{X} \to \mathbb{R}^d$ s.t. for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, $\phi(\mathbf{x})^\top \phi(\mathbf{y}) = (1 + \mathbf{x}^\top \mathbf{y})^2$ must use $d \geq 10$ dims. Give a proof if True else give a map using fewer dimensions as a counter example.	

4	If $X, Y \in \mathbb{R}^{3\times3}$ are rank one matrices, then $X + Y$ can never be rank one, no matter what are X, Y. Give a brief proof if True else give a counter example.

Q2. (Informative non-response models) Melbo is studying how one's income level affects one's reluctance to reveal one's income publicly. n people were chosen with incomes $X_1, X_2, \dots, X_n$. Melbo knows that the income levels X_i are distributed as independent standard Gaussian random variables i.e., $X_i \sim \mathcal{N}(0,1)$ for all i (let us interpret positive X_i as higher-than-median income and negative X_i as lower-than-median income). However, not everyone wants to reveal their income. When Melbo conducts the survey, the responses are $Z_1, Z_2, \dots, Z_n$. If the i^{th} person reveals their income, then $Z_i = X_i$ else $Z_i = \phi$. It is known that $\mathbb{P}[Z_i \neq \phi \mid X_i] = \exp\left(-\frac{\alpha^2 X_i^2}{2}\right)$, where $\alpha > 0$ is an unknown parameter to be learnt. **(Total 12 marks)**

1. Is a rich person e.g., $X_i = 100$ more likely or less likely to reveal their income than a person with close-to-median income e.g., $X_j = -0.01$? Give brief justification. *(1+1 = 2 marks)*

2. Is a poor person e.g., $X_i = -10$ more likely or less likely to reveal their income than a person with close-to-median income e.g., $X_j = 0.1$? Give brief justification. *(1+1 = 2 marks)*

3. Derive an expression for $\mathbb{P}[Z_i \neq \phi]$ the prior probability of a person revealing their income. Show steps and give your answer as a function $h(\alpha)$. **Hint**: the density of a Gaussian looks like $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X-\mu)^2}{2\sigma^2}\right)$ and $X_i \sim \mathcal{N}(0,1)$. Also, $\int_{-\infty}^{\infty} \exp\left(-\frac{a^2t^2}{2}\right)\, dt = \sqrt{\frac{2\pi}{a^2}}$. *(4 marks)*

4. Write down an expression for the negative log-likelihood of the form (no derivation needed)

$$\mathcal{L}(\alpha) = -\sum_{i:Z_i \neq \phi} \ln \mathbb{P}[Z_i \neq \phi, X_i] - \sum_{i:Z_i = \phi} \ln \mathbb{P}[Z_i = \phi]$$

Notice that the terms in the first summation involve joint probability. *(2 marks)*

5. Write down an expression for the gradient $\mathcal{L}'(\alpha)$ (no derivation needed). *(2 marks)*

Q3. (Quantile regression) Can we find the k^{th} largest number in a set of n numbers simply by solving an optimization problem?! Turns out it is indeed possible using a trick called quantile regression. For a set of real numbers $x_1 < x_2 < \cdots < x_n$ (sorted in ascending order for sake of simplicity), for any integer $k = 0,1,2,\dots n$, consider the problem $\operatorname{argmin}_{z\in[x_1,x_n]} f_k(z)$, with

$$f_k(z) \stackrel{\text{def}}{=} \left(\frac{k}{n}-1\right)\cdot\sum\nolimits_{x_i<z}(x_i - z) + \frac{k}{n}\cdot\sum\nolimits_{x_i\geq z}(x_i - z)$$

There are no duplicates in $x_1, \dots, x_n$. Assume that an empty sum equals 0.

1. Find a minimizer for $\operatorname{argmin}_{z\in[x_1,x_n]} f_n(z)$ i.e., $k = n$. Show brief derivation. *(1+1=2 marks)*

2. Find a minimizer for $\operatorname{argmin}_{z\in[x_1,x_n]} f_0(z)$ i.e., $k = 0$. Show brief derivation. *(1+1=2 marks)*

5. Let us handle $k \in [1, n-1]$. Show brief derivation that if $x_j < a < b \leq x_{j+1}$, $a \neq b$, then
   a. We have $f_k(a) > f_k(b)$ if $1 \leq j < k$.
   b. We have $f_k(a) < f_k(b)$ if $k < j < n$, we have.
   c. We have $f_k(a) = f_k(b)$ if $j = k$, i.e., for $x_k < a < b \leq x_{k+1}$. *(4+4+4 = 12 marks)*

After establishing a few more results like the ones above (which you do not have to show), we can deduce that any value of $z \in [x_k, x_{k+1})$ is a minimizer of $\operatorname{argmin}_{z\in[x_1,x_n]} f_k(z)$. **(Total 16 marks)**

Q4. (Robust mean estimation) Melbo has got samples $X_1, \ldots, X_n$ from a Gaussian with unknown mean μ but known variance $\sigma = \frac{1}{\sqrt{2\pi}}$ i.e., with density $f(X; \mu) = \exp(-\pi(X-\mu)^2)$. Melbo wishes to estimate μ using these samples but is stuck since some samples were corrupted by Melbo's enemy Oblem. It is not known which samples did Oblem corrupt. Let's use latent variables to solve this problem. For each i, we say $Z_i = 1$ if we think X_i is corrupted else $Z_i = 0$. For any $\mu \in \mathbb{R}$, we are told that $\mathbb{P}[Z_i = 1 \mid \mu] = \eta$, and that $\mathbb{P}[X_i \mid \mu, Z_i = 1] = \epsilon$, and $\mathbb{P}[X_i \mid \mu, Z_i = 0] = f(X_i; \mu)$. Thus, we suspect that Oblem corrupted around η fraction of the samples and we assume that a corrupted sample can take any value with probability ϵ. Assume $\epsilon, \eta < \frac{1}{10}$ and are both known.

1. For a given μ, derive for a rule to find out if $\mathbb{P}[Z_i = 1 \mid X_i, \mu] > \mathbb{P}[Z_i = 0 \mid X_i, \mu]$ or not.

2. Suppose we are given values of $Z_1, \dots, Z_n \in \{0,1\}$. Derive an expression for the MLE estimate

$$\operatorname{argmax}_{\mu \in \mathbb{R}} \prod_{i=1}^{n} \mathbb{P}[X_i \mid \mu, Z_i]$$

Note that this allows us to execute alternating optimization to help Melbo solve the problem even in the presence of corruptions. We can initialize μ (say randomly), then use part 1 to set Z_i values for each i (set $Z_i = 1$ if $\mathbb{P}[Z_i = 1 \mid X_i, \mu] > \mathbb{P}[Z_i = 0 \mid X_i, \mu]$ else set $Z_i = 0$), then use part 2 to update μ given these Z_i values and then repeat the process till convergence. **(5 + 5 = 10 marks)**