CS 771A: Intro to Machine Learning, IIT Kanpur				**Quiz II** (19 Oct 2022)
Name	MELBO			**20 marks**
Roll No	000001	Dept.	AWSM	Page **1** of **2**

Instructions:
1. This question paper contains 1 page (2 sides of paper). Please verify.
2. Write your name, roll number, department above in **block letters neatly with ink**.
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 – such cases will get straight 0 marks.
5. Do not rush to fill in answers. You have enough time to solve this quiz.

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

1	If $A, B \subset \mathbb{R}^2$ are convex sets, then the set $C = \frac{(A+B)}{2} \stackrel{\text{def}}{=} \left\{\frac{(a+b)}{2}, a \in A, b \in B\right\}$ is always convex. Give a brief proof if True else give counter example if False.	T

Consider any two points $c_1, c_2 \in C$ and suppose $c_1 = \frac{(a_1+b_1)}{2}$ and $c_2 = \frac{(a_2+b_2)}{2}$ where $a_1, a_2 \in A$ and $b_1, b_2 \in B$ as per definition of the set C. Then consider the point $\bar{c} = \frac{(c_1+c_2)}{2}$. Plugging in the above values gives us $\bar{c} = \frac{1}{4}(a_1 + a_2 + b_1 + b_2) = \frac{\bar{a}+\bar{b}}{2}$ where $\bar{a} = \frac{(a_1+a_2)}{2}$ and $\bar{b} = \frac{(b_1+b_2)}{2}$. Since A, B are convex, we have $\bar{a} \in A, \bar{b} \in B$. Thus, we get $\bar{c} \in C$ since it can be expressed as an average of a point from A and a point from B. Thus, C is convex.

2	The solution to $\min_{x \in \mathbb{R}} \frac{\lambda}{2} \cdot x^2 + \sum_{i=1}^n \left(x - a^i\right)^2$ approaches the mean of the real numbers $a^1, a^2, \ldots, a^n \in \mathbb{R}$ as $\lambda \to \infty$. Justify by deriving the solution below.	F

By first order optimality, we know that the solution of the optimization problem must satisfy

$$\lambda \cdot x + 2 \cdot \sum_{i=1}^{n} \left(x - a^i\right) = 0 \Rightarrow x = \frac{2}{\lambda + 2n} \sum_{i=1}^{n} a^i$$

Thus, as $\lambda \to \infty$, we have $x \to 0$. Thus, in general the solution does not approach the mean of the numbers $a^1, \ldots, a^n$ unless coincidentally, the mean of the numbers is also 0.

3	For a doubly differentiable fn $f: \mathbb{R} \to \mathbb{R}$, if $f''(x_0) = 0$ at $x_0 \in \mathbb{R}$, then it must be that $f'(x_0) = 0$. Give brief justification if True else give counter example if False.	F

Consider the linear function $f(x) = x$ which satisfies $f''(x) = 0$ but $f'(x) = 1$.

4	Let X, Y be two random variables (may or may not be independent) such that $\text{Var}[X] > 0$ and $\text{Var}[Y] > 0$. Then it may happen that $\text{Var}[X+Y] = 0$. Give an example if your answer is True else give a proof that $\text{Var}[X+Y] > 0$ always.	T

Let X be a standard Rademacher variable i.e., $\mathbb{P}[X = 1] = 0.5 = \mathbb{P}[X = -1]$ and define a new random variable $Y \overset{\text{def}}{=} -X$ i.e., Y always takes the negative of whatever value X takes. Since we have $\mathbb{E}[X] = 0 = \mathbb{E}[Y]$, negating a variable preserves its variance and $X^2 \equiv 1 \equiv Y^2$, we have

$$\text{Var}[X] = \text{Var}[Y] = 1 > 0$$

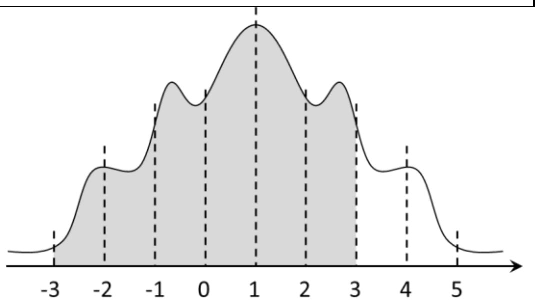
However, by construction, $X + Y \equiv 0$ i.e., $\text{Var}[X+Y] = 0$.

Note that this counter example works only when X, Y are non-independent since if X, Y were independent, then we would instead have

$$\begin{aligned}\text{Var}[X+Y] &= \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2 = \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\ &= \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\mathbb{E}[X]\mathbb{E}[Y] - (\mathbb{E}[X])^2 - (\mathbb{E}[Y])^2 - 2\mathbb{E}[X]\mathbb{E}[Y] \\ &= \text{Var}[X] + \text{Var}[Y]\end{aligned}$$

Thus, if X, Y are independent, then $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] > 0$.

Q2. (Deviant behaviour) Melbo created a new class of $\mathcal{M}$ distributions with mean μ, std σ such that if $X \sim \mathcal{M}(\mu, \sigma^2)$, then for any $c > 0$ we have $\mathbb{P}[\mu - c\sigma \le X \le \mu + c\sigma] = \eta_c$ It is known that $\eta_1 = 0.75, \eta_2 = 0.9, \eta_3 = 0.95, \eta_4 = 0.99$. For $\mu = 1, \sigma = 1$, find out $\mathbb{P}[-3 \le X \le 3]$. Give brief calculations below and the final answer. **(2+2=4 marks)**



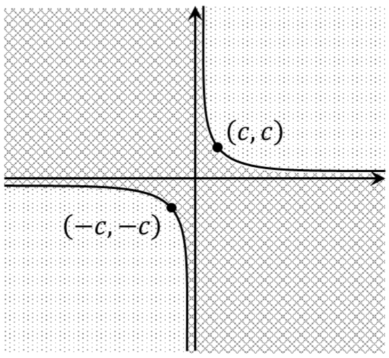
Assuming the distribution to be symmetric, we know that

$$\mathbb{P}[-3 \le X \le 1] = \frac{\eta_4}{2} = 0.495$$

$$\mathbb{P}[1 \le X \le 3] = \frac{\eta_2}{2} = 0.45$$

Thus, we get $\mathbb{P}[-3 \le X \le 1] = \frac{\eta_4 + \eta_2}{2} = 0.945$

Q3. (XOR classifier) For a 2D rectangular hyperbola with equation $xy = c^2$ for $c \in \mathbb{R}$, give a feature map $\phi: \mathbb{R}^2 \to \mathbb{R}^D$ for some $D > 0$ and a corresponding linear classifier $\mathbf{W} \in \mathbb{R}^D$ so that for any $\mathbf{x} \in \mathbb{R}^2$, $\text{sign}(\mathbf{W}^\top \phi(\mathbf{x}))$ takes value $z = -1$ in the light dotted region (see figure on the right) and $z = +1$ in the dark cross-hatched region. We don't care what happens on the boundary. Your map ϕ must **not depend on** c but $\mathbf{W}$ may depend on c. No need to show calculations. Just give the final answers. **(2 + 2 = 4 marks)**



$$\phi(x, y) = [xy, 1]$$

$$\mathbf{W} = [-1, c^2]$$