CS 771A: Intro to Machine Learning, IIT Kanpur				**Midsem Exam** (16 Feb 2020)
Name				**80 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 neatly** with ink **on each page** of this question paper.
3. If you don't write your name and roll number on **all** pages, **pages may get lost** when we unstaple to scan pages
4. Write your final answers neatly **with a blue/black pen**. Pencil marks may get smudged.
5. Don't overwrite/scratch answers especially in MCQ and T/F. We will entertain no requests for leniency.

Q1. Write **T** or **F** for True/False (write **only in the box on the right hand side**) (**3x2 = 6 marks**)

1	The variance of a real-valued (continuous or discrete) random variable must always be a strictly positive number i.e. it can never be zero or negative.	
2	Suppose X, Y are two (not necessarily independent) r.v. with support $\mathbb{R}$. We are told that $\text{Var}[X] \neq 0$ and $\text{Var}[Y] \neq 0$. Then it must be the case that $\text{Var}[X+Y] \neq 0$.	
3	Consider a doubly differentiable function $f: \mathbb{R} \to \mathbb{R}$ that always takes values in the interval $[0,1]$ i.e. $f(x) \in [0,1]$ for all $x \in \mathbb{R}$. Such a function can never be convex.	

Q2. Fill the circle (**don't tick**) next to all the correct options (**many may be correct**).(**2x3=6 marks**)

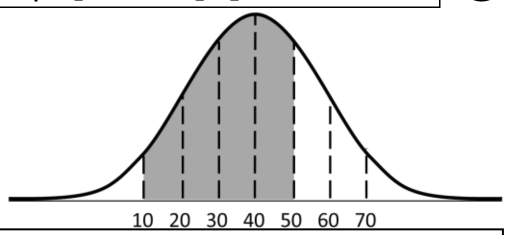
2.1 Let $X \sim \mathcal{N}(0, \sigma^2)$ be a Gaussian r.v.. Let Y be another r.v. defined as $Y = X$ if $X \in [-1,1]$ else $Y = 2X$ if $X \notin [-1,1]$. Suppose we keep increasing σ. Then as $\sigma \to \infty$, which of the following is true?

A	The variance $\text{Var}[X]$ goes up as $\sigma \to \infty$	○
B	The variance $\text{Var}[X]$ goes down as $\sigma \to \infty$	○
C	The PDF $\mathbb{P}\left[Y \mid X \in [-1,1]\right]$ starts looking more and more like $\text{UNIF}[-1,1]$ as $\sigma \to \infty$	○
D	The probability $\mathbb{P}\left[Y \in [-1,1]\right]$ approaches 1 as $\sigma \to \infty$	○

2.2 Suppose X, Y are two independent r.v. Which of the following is always true?

A	$\mathbb{E}[X+Y] = \mathbb{E}X + \mathbb{E}Y$	○
B	$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$	○
C	$\mathbb{E}[X \mid Y] = \mathbb{E}X$	○
D	$\text{Var}[X \mid Y] = \text{Var}[X]$	○

Q3 $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = 40$ and $\sigma = 10$. Approximate $\mathbb{P}[10 \leq X \leq 50]$ using the 68-95-99.7 rule that tells us that the probability that X is within c standard deviations of its mean is approximately 0.68, 0.95 and 0.997 for $c = 1,2,3$ respectively. Explain your answer briefly. **(4 marks)**



Q4 X is a continuous r.v. with support $[0,1]$ and uniform PDF over its support. Define a new discrete r.v. Y with support over $\{0,1,2\}$. Let $0 \leq a \leq b \leq 1$ be two (unknown) constants. Y is defined using X itself as follows: define $Y = \begin{cases} 0 & X \leq a \\ 1 & a < X < b \\ 2 & X \geq b \end{cases}$ (please pay attention to $<$ vs $\leq$). First, write down the PMF for Y. Then write down an expression for $\mathrm{Var}[Y]$. Your PMF and variance expressions must be in terms of the variables a, b and not for specific values of a, b. Then, find a pair of values for a, b for which the r.v. Y has its highest possible variance (give a brief derivation as well). For your chosen value of a, b, write down the value of $\mathbb{E}[X + Y]$ and $\mathrm{Var}[Y]$. **(4+4+4+2+2=16 marks)**

Give PMF for Y here

Give expression for $\mathrm{Var}[Y]$ here

Best value of $a = (\qquad)$ $b = (\qquad)$

Give derivation for best values of a,b here

$\mathbb{E}[X + Y] = (\qquad)$ $\mathrm{Var}[Y] = (\qquad)$

Q5 Let $a^1, \ldots, a^n \in \mathbb{R}$ and $\lambda > 0$ be constants that are given to us. **(4+4+10+6 = 24 marks)**

5.1 Solve the L_2 regularized problem (give brief derivation) $\min_x \frac{\lambda}{2} \cdot x^2 + \frac{1}{2n} \cdot \sum_{i=1}^n \left(x - a^i\right)^2$

5.2 We will now solve $\min_x \lambda \cdot |x| + \frac{1}{2n} \cdot \sum_{i=1}^n \left(x - a^i\right)^2$ the L_1 regularized problem. We rewrite the optimization problem as shown on the right-hand side. Write down the Lagrangian of this problem by introducing dual variables for the constraints.

$$\begin{aligned} \min_{x,c} \; & \lambda \cdot c + \frac{1}{2n} \cdot \sum_{i=1}^n \left(x - a^i\right)^2 \\ & x \leq c \\ \text{s.t.} \quad & x \geq -c \\ & c \geq 0 \end{aligned}$$

5.3 Using the Lagrangian, create and simplify the dual problem (show brief derivation). If simplified properly, the dual problem should involve only two real valued variables. Try to simplify the dual problem as much as you can (otherwise the next part may be more difficult for you).

5.4 Propose a coordinate ascent/maximization (or descent/minimization if you are writing your dual as a minimization problem) method to solve your simplified dual. Use cyclic coordinate selection and random initialization for simplicity. Give precise expressions in your pseudocode (not vague statements) on how you would process a chosen coordinate taking care of constraints etc.

Give solution to 5.1 here

Give solution to 5.2 here

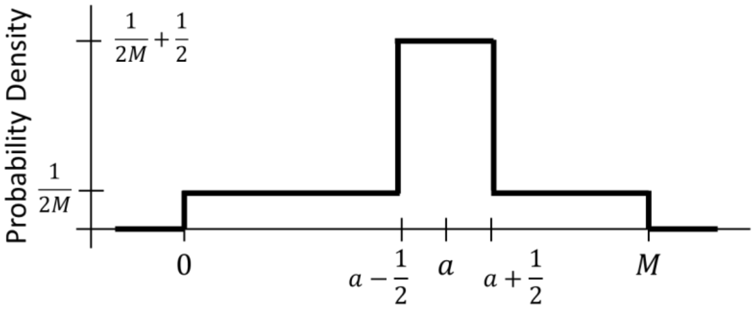
Give solution to 5.3 here

Give solution to 5.4 here

CS 771A: Intro to Machine Learning, IIT Kanpur				**Midsem Exam** (16 Feb 2020)
Name				**80 marks**
Roll No		Dept.		

Q6 Let $M \gg 1$ be some large constant. For $a \in [0.5, M-0.5]$, define an *almost-uniform* (AU) distribution $\tilde{\mathcal{U}}$ with PDF described below. We are given data $x^1, \ldots, x^n \in [0, M]$. We wish to fit a mixture of two AU distributions $\tilde{\mathcal{U}}(x; a^1)$ and $\tilde{\mathcal{U}}(x; a^2)$ for $a^1, a^2 \in [0.5, M-0.5]$ to the data.

$$\mathbb{P}[x \mid a] = \tilde{\mathcal{U}}(x; a) \triangleq \begin{cases} 0 & x \notin [0, M] \\ \frac{1}{2M} & x \in \left[0, a - \frac{1}{2}\right) \\ \frac{1}{2M} + \frac{1}{2} & x \in \left[a - \frac{1}{2}, a + \frac{1}{2}\right] \\ \frac{1}{2M} & x \in \left(a + \frac{1}{2}, M\right] \end{cases}$$



Let $z^i \in \{1,2\}$ be latent variables denoting which distribution generated which data point. Use the data likelihood function $\mathbb{P}\left[x^i \mid z^i, a^1, a^2\right] = \tilde{\mathcal{U}}\left(x^i; a^{(z^i)}\right)$. Expressions in your answers may contain unspecified normalization constants. Give only brief derivations. **(8+10+6=24 marks)**

6.1 Assuming $\mathbb{P}\left[z^i = c \mid a^1, a^2\right] = \mathbb{P}\left[z^i = c\right] = 0.5$ for $c \in \{1,2\}$ (i.e. uniform prior on z^i), derive an expression for $\mathbb{P}\left[z^i = 1 \mid x^i, a^1, a^2\right]$. Using this, show how to calculate the MAP estimate for the latent variables i.e. $\text{argmax}_{c \in \{1,2\}} \mathbb{P}\left[z^i = c \mid x^i, a^1, a^2\right]$. Break any ties in any way you like.

6.2 Show how to calculate the MAP estimate for the model i.e. $\underset{a^1, a^2 \in [0.5, M-0.5]}{\arg\max} \mathbb{P}[a^1, a^2 \mid \mathbf{X}, \mathbf{z}]$ (note the shorthand $\mathbf{X} = [x^1, \ldots, x^n] \in [0, M]^n, \mathbf{z} = [z^1, \ldots, z^n] \in \{1,2\}^n$). You are allowed to use $\texttt{k-nn}$ and $\texttt{r-nn}$ as library function calls in your solution. For any $x \in [0, M], k \in \mathbb{N}, r \in \mathbb{R}_+$, $\texttt{k-nn}(x, k)$ returns the k nearest neighbors of x in the dataset $\mathbf{X}$ whereas $\texttt{r-nn}(x, r)$ returns all points in the dataset $\mathbf{X}$ which are a distance r or less from x. Both $\texttt{k-nn}$ and $\texttt{r-nn}$ use the Euclidean distance.

6.3 Using the above, give pseudocode (as we do in lecture slides with sufficient algo details and not necessarily Python code) for a k-means style alternating optimization algorithm (and not "soft" k-means or EM) for estimating the model $a^1, a^2 \in [0.5, M-0.5]$. Use random initialization for sake of simplicity. Give precise update expressions in pseudocode and not just vague statements.

Give solution to 6.1 here

Give solution to 6.2 here

Give solution to 6.3 here

- -- END OF EXAM - - - - -- -