CS 771A: Intro to Machine Learning, IIT Kanpur				**Quiz I** (22 Jan 2020)
Name	SAMPLE SOLUTIONS			**30 marks** Page **1** of **2**
Roll No		Dept.		

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. We will entertain no requests for leniency.
5. Do not rush to fill in answers. You have enough time to solve this quiz.

Q1. Write **T** or **F** for True/False (write **only in the box on the right hand side**) **(5x2=10 marks)**

1	If a set $\mathcal{C} \subset \mathbb{R}^2$ is convex, then all subsets of $\mathcal{C}$ must be convex sets as well i.e. if $\mathcal{C}' \subseteq \mathcal{C}$ then $\mathcal{C}'$ must be convex too	F
2	Let $f: \mathbb{R} \to \mathbb{R}$ be a doubly differentiable function (i.e. first and second derivatives exist). If $f'(x^0) > 0$ at $x^0 \in \mathbb{R}$, then it may be possible that $f''(x^0) < 0$	T
3	For a binary classification problem with feature vectors in $\mathbb{R}^5$, a linear model will in general have smaller size than an LwP model (if using one prototype per class)	T
4	Suppose $f: \mathbb{R} \to \mathbb{R}$ is a differentiable function but neither convex nor concave. Then it must be the case that f has strictly more than one local optima	F
5	There can exist a training dataset for a binary classification problem on which the 1NN (one nearest neighbour) algorithm has a linear decision boundary	T

Q2. (Squared Hinge Loss) Let $\mathbf{a} \in \mathbb{R}^d$ and $b \in \mathbb{R}$ be constants. For $\mathbf{x} \in \mathbb{R}^d$, define the function $f(\mathbf{x}) = ([1 - b \cdot \mathbf{a}^\top \mathbf{x}]_+)^2$. Find $\nabla f(\mathbf{x})$ and briefly show all major steps in your derivation **(5 marks)**

The function $g(t) = ([1-t]_+)^2$ is actually differentiable. The function is clearly differentiable for $t \neq 1$ with $g'(t) = \begin{cases} 0 & t > 1 \\ 2(t-1) & t < 1 \end{cases}$. However, it is clear that $\lim_{t \to 1^+} g'(t) = \lim_{t \to 1^-} g'(t) = 0$ which shows that $g(t)$ is differentiable. Now, we simply have $f(\mathbf{x}) = g(b \cdot \mathbf{a}^\top \mathbf{x})$ and thus, on applying chain rule we get $\nabla f(\mathbf{x}) = g'(b \cdot \mathbf{a}^\top \mathbf{x}) \cdot b \cdot \mathbf{a} = \begin{cases} \mathbf{0} & b \cdot \mathbf{a}^\top \mathbf{x} \geq 1 \\ 2b(b \cdot \mathbf{a}^\top \mathbf{x} - 1) \cdot \mathbf{a} & b \cdot \mathbf{a}^\top \mathbf{x} < 1 \end{cases}$.
Note that the $\mathbf{0}$ above is the zero vector and not the zero real number.

Q3. (Model Exfiltration) Ms M has a secret function $f: \mathbb{R}^4 \to \mathbb{R}$. We know that f thresholds one of the 4 coordinates of the input at a value i.e. for all $\mathbf{x} \in \mathbb{R}^4$, $f(\mathbf{x}) = \mathbf{x}_j - c$ where $j \in \{1,2,3,4\}$ and $c \in \mathbb{R}$. E.g. if $j = 2, c = 1.5$, then for $\mathbf{x} = (1,5,2,2)$, $f(\mathbf{x}) = 5 - 1.5 = 3.5$. I want to steal Ms M's model and find out what value of j, c is she using. I can send Ms M any number of inputs $\mathbf{x}^1, \mathbf{x}^2, \ldots \in \mathbb{R}^4$ and she will return $f(\mathbf{x}^1), f(\mathbf{x}^2), \ldots \in \mathbb{R}$ back to me. Design an algorithm below that asks Ms M function values on one or more 4D vectors and uses her responses to find the value of j and c being used. You must give explicit descriptions of the 4D vectors you are querying Ms M (e.g. you may say that you wish to query only three vectors $(1,-0.5,1,-3)$, $(8,-0.1,1,5)$ and $(9,1,5,6)$). The fewer vectors you query Ms M to correctly find j, c, the more marks you will get. **(10 marks)**

Step 1: Query $\mathbf{x}^1 = (0,0,0,0)$. Note that $f(\mathbf{x}^1) = -c$ so $c = -f(\mathbf{x}^1)$

Step 2: Query $\mathbf{x}^2 = (1,2,3,4)$. Note that $f(\mathbf{x}^2) = \mathbf{x}_j^2 - c = \mathbf{x}_j^2 + f(\mathbf{x}^1)$ so $\mathbf{x}_j^2 = f(\mathbf{x}^2) - f(\mathbf{x}^1)$
Since we have deliberately set $\mathbf{x}_i^2 = i$ for all $i \in \{1,2,3,4\}$, we get $j = f(\mathbf{x}^2) - f(\mathbf{x}^1)$

Step 3: Output $c = -f(\mathbf{x}^1), j = f(\mathbf{x}^2) - f(\mathbf{x}^1)$

Note that the only thing required in step 2 is for $\mathbf{x}^2$ to have different values for all its four coordinates. This helps us figure out which coordinate is being thresholded. Setting $\mathbf{x}_i^2 = i$ merely makes the algorithm more aesthetically pleasing ☺

Also note that the above scheme uses two queries to figure out the model completely. This is in general optimal i.e. there is no way an algorithm can figure out both j, c using just one query. To see why, suppose an algorithm makes a single query on a vector $\mathbf{x} = (p, q, r, s)$. Upon receiving the response, such an algorithm will still be unsure whether $j = 1$ and $c = p - f(\mathbf{x})$ or whether $j = 2$ and $c = q - f(\mathbf{x})$ or $j = 3$ and $c = r - f(\mathbf{x})$ or $j = 4$ and $c = s - f(\mathbf{x})$. The algorithm will necessarily need to make another query to zero-in on one of the above 4 cases.

Q4. (Placement Woes) Ms M wants to set up a shop at a point in $\mathbb{R}^2$. The consumer density is such that if she opens shop at a point $(x, y) \in \mathbb{R}^2$, her daily income will be $2x + 4y$. However, since her supplier is situated at $(2, -1)$, she will incur a daily cost of $(x-2)^2 + (y+1)^2$ for transporting goods from her supplier to her shop. Where should Ms M open her shop to get the maximum daily profit and what is that maximum value of profit? Give a brief derivation below. **(5 marks)**

We cast the above problem as an optimization problem. Let $\mathbf{a} = (2,4), \mathbf{b} = (2,-1) \in \mathbb{R}^2$. We wish to minimize $f(\mathbf{x}) = \|\mathbf{x} - \mathbf{b}\|_2^2 - \mathbf{a}^\top \mathbf{x}$ (minimizing loss is the same as maximizing profit). We have $\nabla f(\mathbf{x}) = 2(\mathbf{x} - \mathbf{b}) - \mathbf{a}$ and $\nabla^2 f(\mathbf{x}) = 2I \succ 0$ i.e. f is a convex differentiable function. Using first order optimality to get the global minimum gives us $\mathbf{x}^{\text{opt}} = \mathbf{b} + \mathbf{a}/2 = (3,1)$ and the profit earned by opening up a shot at $\mathbf{x}^{\text{opt}}$ is $-f(\mathbf{x}^{\text{opt}}) = 5$ units.