CS 771A: Introduction to Machine Learning				**Quiz 2** (30 Aug 2019)
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**) **(8x2=16 marks)**

1	When minimizing a convex function f using (sub)gradient descent (without any constraints), it does not matter what step lengths we choose since f is convex	F
2	Consider the function $f: \mathbb{R} \to \mathbb{R}$ defined as $f(x) = 2x^2 + x + 2$, the subdifferential of f at $x = -1$ i.e. $\partial f(-1)$ is the set $\{-3, 3\}$	F
3	Executing one step of mini-batch stochastic gradient descent with large batch size is usually more expensive that executing one step of stochastic gradient descent	T
4	The Hessian of a function $f: \mathbb{R}^2 \to \mathbb{R}$ is always a 2×2 PSD matrix	F
5	If $f, g: \mathbb{R} \to \mathbb{R}$ are two non-differentiable functions, then the function $f + g$ will always be non-differentiable as well	F
6	k-means++ is a more powerful algorithm than k-means since k-means++ uses the Mahalanobis distance instead of the Euclidean distance to cluster data points	F
7	Suppose X is a random variable such that $x \geq 1$ for all $x \in S_X$ where S_X is the support of X. Then the variance of X must be greater than 1 too i.e. $\mathbb{V}[X] \geq 1$	F
8	Suppose we define Y as the indicator variable of an event A i.e. $Y = \mathbb{I}\{A\}$. Then $\mathbb{E}[Y]$ can never be strictly negative i.e. we must always have $\mathbb{E}[Y] \geq 0$	T

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

2.1 Suppose X, Y are two random variables (r.v. for short) such that $X + Y = 4$ i.e. for all outcomes $\omega \in \Omega$, we have $X(\omega) + Y(\omega) = 4$. Which all of the following statements is true?

A	We must have $\mathbb{E}[X + Y] = 4$	●
B	If we are told that $\mathbb{E}[X] = 5$, then $\mathbb{E}[Y]$ cannot be positive	●
C	We cannot have $\mathbb{V}[X + Y] = 0$ i.e. we must have $\mathbb{V}[X + Y] > 0$	○
D	If X is a constant r.v. then Y must be a constant r.v. too	●

2.2 Let Z denote a random variable with support $S_Z = \{0, 2\}$. Then which all of the given statements is true about Z?

A	We must always have $\mathbb{E}[Z] \leq 1$	○
B	We must always have $\mathbb{V}[Z] \leq 1$	●
C	We must always have $\mathbb{E}[Z] \leq 2$	●
D	We must always have $\mathbb{V}[Z] \leq 2$	●

Q3 Consider the optimization problem to the right. Write down the Lagrangian of the problem. Then write down the dual of the problem. Eliminate the primal variable and write down the simplified dual problem with the primal variable eliminated. Then solve the dual problem and write down the optimal values of the primal and dual variables that you have obtained.

$$\min_{x\in\mathbb{R}} \frac{1}{2}x^2 \quad \text{s.t.} \quad x \geq 2$$

(1+1+1+2 = 5 marks)

If we introduce a dual variable α, the Lagrangian is $\mathcal{L}(x,\alpha) = \frac{1}{2}x^2 + \alpha(2-x)$

The dual problem is $\max\limits_{\alpha\geq 0}\min\limits_{x}\frac{1}{2}x^2 + \alpha(2-x)$

Using first order optimality gives us $x - \alpha = 0$ on in other words, $x = \alpha$. Substituting this into the dual problem gives us $\max\limits_{\alpha\geq 0} 2\alpha - \frac{1}{2}\alpha^2$ as the simplified dual problem.

If suppose we ignore the constraint $\alpha \geq 0$ for a moment, then applying the first order optimality condition tells us that the function $2\alpha - \frac{1}{2}\alpha^2$ achieves its maximum where $2 - \alpha = 0$ or in other words at $\alpha = 2$. However, this point does satisfy $\alpha \geq 0$ which means $\alpha = 2$ is the maximum for $2\alpha - \frac{1}{2}\alpha^2$ even with the constraint $\alpha \geq 0$ and thus, the solution of the dual problem . Using $x = \alpha$ and strong duality tells us that $x = 2$ is the solution to the primal problem.

Q4. Let $\mathbf{a} \in \mathbb{R}^d$ be a constant vector and $b \in \mathbb{R}$ be a constant scalar. Let $f: \mathbb{R}^d \to \mathbb{R}$ be a function defined as $f(\mathbf{x}) = \max\{\mathbf{a}^\top\mathbf{x} + b, 0\}$. Find an expression for subdifferential of f at any $\mathbf{x} \in \mathbb{R}^d$. Show all the main steps of your derivation briefly in the space provided. **(3 marks)**

We have $f(\mathbf{x}) = \max\{g(\mathbf{x}), h(\mathbf{x})\}$ where $g(\mathbf{x}) = \mathbf{a}^\top\mathbf{x} + b$ and $h(\mathbf{x}) = 0$. Note that both $g(\mathbf{x})$ and $h(\mathbf{x})$ are differentiable functions. Thus, applying the max rule for subgradients gives us

$$\begin{cases} \mathbf{a} & \text{if } \mathbf{a}^\top\mathbf{x} + b > 0 \\ \mathbf{0} & \text{if } \mathbf{a}^\top\mathbf{x} + b < 0 \\ \lambda\cdot\mathbf{a} & \text{if } \mathbf{a}^\top\mathbf{x} + b = 0 \end{cases}$$

where $\lambda \in [0,1]$ and $\mathbf{0} \in \mathbb{R}^d$ is the d-dimensional all zeros vector. Note that in the second case, the subgradient is the zero vector, not the number/ scalar zero.

---------------------------------- END OF QUIZ ----------------------------------