CS 771A: Intro to Machine Learning, IIT Kanpur				**Quiz II** (20 Mar 2024)
Name				**20 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 – such cases will get straight 0 marks.
5. Do not rush to fill in answers. You have enough time to solve this quiz.

(Noise to Regularize) The underlying principle behind the deep learning technique *dropout* is that adding noise to data can prevent models from overfitting. Let us derive this fact formally. **Q1.** Let $\boldsymbol{\epsilon} \in \{-1, +1\}^D$ be a D-dim Rademacher vector with coordinates chosen i.i.d. $\epsilon_j = 1$ or -1 uniformly randomly. Find the following (no derivation) Note: $j, k \in [D], j \neq k$. **(6 x 1 = 6 marks)**

$\mathbb{E}[\epsilon_j + \epsilon_k] =$	$\mathrm{Var}[\epsilon_j + \epsilon_k] =$
$\mathbb{E}[\epsilon_j \epsilon_k] =$	$\mathrm{Var}[\epsilon_j \epsilon_k] =$
$\mathbb{E}[\epsilon_j / \epsilon_k] =$	$\mathrm{Var}[\epsilon_j / \epsilon_k] =$

Q2. Let $y, \lambda \in \mathbb{R}$ and $\mathbf{w}, \mathbf{x} \in \mathbb{R}^D$ be constants and $\boldsymbol{\epsilon} \in \{-1, +1\}^D$ be a Rademacher vector sampled independently of $y, \lambda, \mathbf{w}, \mathbf{x}$. Obtain a simplified expression (expectation is over the choice of $\boldsymbol{\epsilon}$ only). Give brief derivation. Your expression should not contain any ϵ_j terms. **(2 + 4 = 6 marks)**

Write final expression in the box

$\mathbb{E}_{\boldsymbol{\epsilon}}\left[\left(y - \mathbf{w}^\top(\mathbf{x} + \lambda \cdot \boldsymbol{\epsilon})\right)^2\right] =$

Brief derivation for simplification

Q3. We have N datapoints $(\mathbf{x}^n, y^n) \in \mathbb{R}^D \times \mathbb{R}, n \in [N], \lambda \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^D$ all of which can be treated as constants. We also sample N Rademacher vectors $\boldsymbol{\epsilon}^n \in \{-1, +1\}^D, n \in [N]$ i.i.d. of each other as well as independent of the datapoints and $\lambda, \mathbf{w}$. Expectation is over the choice of $\{\boldsymbol{\epsilon}^n, n \in [N]\}$ only. Write down a simplified expression for the following (no derivation needed). **(2 marks)**

$$\mathbb{E}_{\{\boldsymbol{\epsilon}^n\}}\left[\sum_{n\in[N]}\left(y^n - \mathbf{w}^\top(\mathbf{x}^n + \lambda \cdot \boldsymbol{\epsilon}^n)\right)^2\right] =$$

Q4 (IPL Intrigue). Melbo is a big IPL fan and is trying to analyse the performance of MI vs CSK on various kinds of pitches. Let M be the event that MI won a MI-vs-CSK match and C be the event that CSK won a MI-vs-CSK match. There are 3 kinds of pitches $F =$ flat, $G =$ green, $D =$ dusty. A total of 24 matches were played between MI and CSK, 1/4$^{\text{th}}$ of which were on green pitches and 1/3$^{\text{rd}}$ on flat pitches. MI won 6 of the matches played on flat pitches. Both MI and CSK won equal number of matches played on green pitches i.e., $\mathbb{P}[M \mid G] = \mathbb{P}[C \mid G]$. Also, both flat and dusty pitches have been equally favourable for MI in that $\mathbb{P}[F \mid M] = \mathbb{P}[D \mid M]$. Find out the following quantities as fractions or decimals (no derivations needed). *Hint: either use Bayes rule or fill-up a 2 x 3 matrix showing which team won how many matches on what kind of pitch*. **(6 x 1 = 6 marks)**

$\mathbb{P}[F \mid M] =$	$\mathbb{P}[F \mid C] =$
$\mathbb{P}[G \mid M] =$	$\mathbb{P}[G \mid C] =$
$\mathbb{P}[D \mid M] =$	$\mathbb{P}[D \mid C] =$

Anything written here will not be graded