Quiz 2

● Graded

Student

LAVESH GUPTA

Total Points

20 / 20 pts

Question 1

N2R1

6 / 6 pts

+ 0 pts Give marks directly out of 6 using points adjustment. 1 mark per correct answer. No partial marking for incorrect sign etc.

+ 0 pts Completely wrong or else unanswered

+ 6 pts Point adjustment

Question 2

N2R2

6 / 6 pts

− 1 pt Minor mistakes such as forgetting a constant term or wrong sign

✔ **+ 2 pts** Valid expression for the expectation, possibly with minor mistakes. Completing the squares is not essential.

✔ **+ 4 pts** Brief derivation using linearity of expectation, possibly with mistakes

− 1 pt Minor mistakes in derivation e.g. wrong sign or wrong multiplier

− 2 pts Major mistakes in derivation e.g. not applying linearity of expectation correctly or other major errors

+ 0 pts Totally wrong or else unanswered

Question 3

N2R3

2 / 2 pts

✔ **+ 2 pts** Valid expression for the expectation, possibly with minor mistakes.

− 1 pt Minor mistakes e.g. wrong sign or missing terms or missing factor of N in the regularizer.

+ 0 pts Completely wrong or else unanswered

Question 4

IPL Intrigue

6 / 6 pts

+ 0 pts Give marks directly out of 6 using points adjustment. 1 mark per correct answer. No partial marking for incorrect sign etc.

+ 0 pts Completely wrong or else unanswered

✔ **+ 6 pts** All correct

CS 771A: Intro to Machine Learning, IIT Kanpur			**Quiz II** (20 Mar 2024)
Name	LAVESH GUPTA		**20 marks**
Roll No	210562	Dept. ELECTRICAL	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.

(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)**

$\mathbb{E}[\epsilon_j + \epsilon_k] = 0$	$\mathrm{Var}[\epsilon_j + \epsilon_k] = 2$
$\mathbb{E}[\epsilon_j \epsilon_k] = 0$	$\mathrm{Var}[\epsilon_j \epsilon_k] = 1$
$\mathbb{E}[\epsilon_j / \epsilon_k] = 0$	$\mathrm{Var}[\epsilon_j / \epsilon_k] = 1$

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

$$\mathbb{E}_\epsilon\left[\left(y - \mathbf{w}^\top(\mathbf{x} + \lambda \cdot \boldsymbol{\epsilon})\right)^2\right] = \boxed{\begin{aligned} &E[(y - W^Tx)^2] + \lambda^2 \|W\|_2^2 \\ &= (y - W^Tx)^2 + \lambda^2 \|W\|_2^2 \end{aligned}}$$

Brief derivation for simplification

$$E_\epsilon[(y - W^Tx)^2 + \lambda(W^T\epsilon)^2 - 2y\lambda W^T\epsilon]$$

$$= E[(y - W^Tx)^2] + \lambda^2 E[(W^T\epsilon)^2] - 2y\lambda E[W^T\epsilon] \quad \text{—} \circledast$$

$$\cdot\ E[(W^T\epsilon)^2] = E[\textstyle\sum w_i^2 \epsilon_i^2 + 2\sum w_i w_j \epsilon_i \epsilon_j] \qquad \epsilon_i^2 = 1$$

$$= \cancel{E[\|W\|_2^2]}\ \|W\|_2^2 + 0 = \|W\|_2^2 \qquad E(\epsilon_i\epsilon_j) = 0$$

$$= \cancel{E[\|W\|_2^2]}\ \|W\|_2^2 \qquad E(\epsilon_i) = 0$$

$$\cdot\ E[W^T\epsilon] = E[\textstyle\sum w_i \epsilon_i] = 0$$

Continuing, Put values in $\circledast$

$$= E[(y - W^Tx)^2] + \lambda^2 \cancel{E[\|W\|_2^2]}\ \|W\|_2^2$$

$$= (y - W^Tx)^2 + \lambda^2 \|W\|^2$$

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 $\{\epsilon^n, n \in [N]\}$ only. Write down a simplified expression for the following (no derivation needed). **(2 marks)**

$$\mathbb{E}_{\{\epsilon^n\}}\left[\sum_{n\in[N]} \left(y^n - \mathbf{w}^\top(\mathbf{x}^n + \lambda \cdot \boldsymbol{\epsilon}^n)\right)^2\right] = \left\{\sum_{n\in[N]} E\left[(y^n - w^T x^n)^2\right]\right\} + \lambda . N . \|W\|_2^2$$

$$= \left\{\sum_{n\in[N]} (y^n - w^T x^n)^2\right\} + \lambda . N . \|W\|_2^2$$

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, 1/4th of which were on green pitches and 1/3rd 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)**

$\mathbb{P}[F \mid M] =$ 2/5	$\mathbb{P}[F \mid C] =$ 2/9
$\mathbb{P}[G \mid M] =$ 1/5	$\mathbb{P}[G \mid C] =$ 3/9
$\mathbb{P}[D \mid M] =$ 2/5	$\mathbb{P}[D \mid C] =$ 4/9

Anything written here will not be graded

ROUGH WORK