

Quiz 1 for **Machine, Data and Learning**

Date of Exam: 08/02/2021

Max Time: 35 minutes

General Instructions:

- This is an open book (and internet) exam and should be done individually. You can use calculators for computation purposes.
- Clarifications will not be provided during the exam. In case of doubt, please make reasonable assumptions and specify the assumptions clearly. You may lose points if there was no reason to make those assumptions.
- All your answers should be handwritten by you with a pen. You will not receive points for answers that are not handwritten.
- Please write your roll number on the top of each page.
- Please scan and upload your answer script to Moodle. Please name your file xxxxxxxxxxxquiz1.pdf where xxxxxxxxxxx would be your roll number.
- Please keep track of time and make sure to submit your answer script before the portal closes. Please double check that your answer script is clear and legible for grading purposes before you submit.

Note: The questions below are worth >40 points but maximum achievable points are 40 i.e., if your score > 40, will be set to 40.

1. What is your favorite AI/ML (Artificial Intelligence/Machine Learning) application ? Please present what is the AI or ML technique it uses (please be specific) and explain in a couple of sentences why you think that particular technique was needed for this application (i.e. why is the technique mentioned well suited) ? -- (1+2+2) points
2. Consider a function $y = f(x) = ax^2 + bx + c$ where a is last digit of your roll number, b is your team number in assignment 1 and c is last 2 digits of your roll number.

[Regd. b, if you do not have a team number in assignment 1 for any reason please use a random 4 digit number where all the digits are unique and non-zero. If it is verified that you do have a team number, you will receive 0 points].

- a. Please write down the values for a, b and c. -- (2) points
- b. If you were to model $y' = f'(x)$ by replacing the term a with $2*a$ in $f(x)$ and model $y'' = f''(x)$ by replacing the term b with $2*b$ in $f(x)$, which function among f' and f'' would have less distance w.r.t. $f(x)$ when you consider x from 0 to 5. To compute distance between two functions $f(x)$ and $g(x)$ with x from i to j , please use:

$$\sqrt{(f(i) - g(i))^2 + (f(i+1) - g(i+1))^2 + \dots + (f(j) - g(j))^2}. \text{ -- (3) points}$$

- c. Let's say you are given a budget B which you can use to change y' (i.e., $f'(x)$). You need to make the changes by subtracting suitable integer values $v_i \geq 0$, from each possible value of y' (generated using x from 0 to 5), where the sum of v_i 's (i.e., sum of v_0, v_1, v_2, v_3, v_4 and v_5 denoted s_{vi}) is upto B (i.e., $0 \leq s_{vi} \leq B$), with the aim to reduce the distance between $f'(x)$ and $f(x)$ (as computed in (b)). What is the best set of v_i 's to use so the distance is minimized along with explanation for why mathematically this is the best way to do ? Please set the budget B to $(3*b)$. -- (6) points

3. Please explain what dimensionality reduction means and how it is useful in the context of ML ? Let's suppose that your mother tongue M has 100 alphabets and you recently learnt a new language E with 26 alphabets (e.g., English). As a non-native speaker of E you think in M and then translate the same into E to speak in E . Given that you are expressing the same thoughts in lower number of alphabets, would this mean you are performing a dimensionality reduction here ? Please explain your reasoning in detail ? -- (2+4) points
4. You are given the following data set (x,y) (where $y = f(x)$): $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5), (x_6, y_6)$, where

x_1 = Last 3 digits of your roll number [if last 3 digits are of form 0xx or 00x or 000 please use 6xx or 32x or 654]
 x_2 = Last 2 digits of your roll number [if last 2 digits are of form 0x or 00, please use 7x or 54]
 x_3 = Last 1 digit of your roll number
 x_4 = Sum of the digits of x_1 [e.g., if $x_1 = 396$, $x_4 = 3+9+6 = 18$]
 x_5 = Sum of the digits of x_2
 $x_6 = (x_4 * x_5)$
 $y_1 = x_1, y_2 = x_2, y_3 = x_3, y_4 = x_4, y_5 = x_5, y_6 = x_6$

You performed 3-fold cross validation to obtain 3 different realizations of the model, say MR1, MR2 and MR3. You obtained the following model fits using MR1, MR2 and MR3 for each of the points $\{x_1, x_2, x_3, x_4, x_5, x_6\}$:

$\{y_6, y_5, y_4, y_3, y_2, y_1\}$ using MR1, $\{y_3, y_2, y_1, y_6, y_5, y_4\}$ using MR2 and $\{y_1, y_2, y_3, y_4, y_5, y_6\}$ using MR3. Given all the information presented, please

answer the questions below along with all the steps and computations involved. [If there is anything unclear, please make suitable interpretation (in lines of assignment 1 of the course), present the interpretation made and reason for it so we know it is reasonable and solve the question.]

- Please write down the values for x_1, x_2, x_3, x_4, x_5 and x_6 . -- (2) points
- Please write down the formula for bias ? Briefly explain what the formula represents ? -- (1+1) points
- Please compute bias for the points $\{x_1, x_2, x_3, x_4, x_5, x_6\}$. -- (6) points
- Please write down formula for variance ? Briefly explain what the formula represents ? -- (1+1) points
- Please compute variance for the points $\{x_1, x_2, x_3, x_4, x_5, x_6\}$. -- (6) points
- If there is no noise present in the data set, please compute the MSE (Mean Square Error) for the points $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ using results from the above questions. Please write down the formula being used before presenting the computations. -- (3) points