Final Exam for **Machine, Data and Learning**

Date of Exam: 19/04/2021

Expected time and penalty structure: Details shared in Moodle post

Maximum obtainable marks: 50 points [out of 53 points]

General Instructions:

- This is an open book (and internet) exam and should be done individually. You can use calculator 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 name your file xxxxxxxxxxfinal.pdf where xxxxxxxxxx would be your roll number.
- Please double check that your answer script is clear and legible for grading purposes before you submit. Please follow instructions shared on Moodle post regarding submission details and penalty related issues.

1. Please answer the following about Bayes nets:

a. What are the components of a Bayes net ? Please specify along with one line explanation of each component ? -- (2) points

b. Given two different Bayes nets built using the same set of nodes that describe a scenario, how do you say one Bayes net is better than the other ? Please provide two points atleast along with reasons for why these two points matter ? -- (3) points

c. If a CPT table provides the probability for P(A|B,C), what can you say about the relationship between A-B, B-C and C-A ? -- (3) points

2. You are given the following data set (x,y) (where y = f(x)): (x1, y1), (x2,y2), (x3,y3), (x4,y4), (x5,y5), (x6,y6), where

x1 = 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 634]

x2 = Last 2 digits of your roll number [if last 2 digits are of form 0x or 00, please use 6x or 34]

x3 = Last 1 digit of your roll number

x4 = Sum of the digits of x1 [e.g., if x1 = 396, x4 = 3+9+6 = 18]

x5 = Sum of the digits of x2

x6 = (x4 * x5)

y1 = x1, y2 = x2, y3 = x3, y4 = x4, y5 = x5, y6 = x6

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 {x1,x2,x3,x4,x5,x6}: {y6,y5,y4,y1,y2,y3} using MR1, {y2,y1,y4,y3,y6,y5} using MR2 and {y1,y2,y3,y6,y5,y4} 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.]

a. Please write down the values for x1, x2, x3, x4, x5 and x6. -- (2) points
b. Please write down the formula for bias ? Briefly explain what the formula represents ? -- (1+1) points
c. Please compute bias for the points {x1,x2,x3,x4,x5,x6}. -- (6) points
d. Please write down formula for variance ? Briefly explain what the formula represents ? -- (1+1) points
e. Please compute variance for the points {x1,x2,x3,x4,x5,x6} . -- (6) points
f. If there is no noise present in the data set, please compute the MSE (Mean Square Error) for the points {x1,x2,x3,x4,x5,x6} using results from the above questions. Please write down the formula being used before presenting the computations. -- (3) points

3. An agent is in state (x,y) in an MDP. The agent can take actions N (North), S (South), E (East) and W (West) leading to states (x-1, y), (x+1, y), (x, y+1) and (x, y-1) respectively. Action is stochastic i.e., whenever action is taken it goes to the intended state with probability of 0.6 and in perpendicular directions with 0.2 probability each. Each step (or action taken) incurs a cost i.e. receives reward of -1. The discount factor is 0.xy, where x and y are last 2 digits of your roll number i.e., y is the last digit [if last 2 digits are of form 0x or 00, please use 2x or 53]. The current utility values (i.e., at time step t) for each of the 4 states (in the order mentioned earlier) are 4, 5, 4 and 1 (i.e., U_t(state)). What is the best action for the agent to take from (x,y) - please show (all the) utility computations for each of the 4 actions and then mention the best action as outcome of these computations. -- (7) points

4. With the recent rise of positive corona cases, the World Health Organization (WHO) decided to install mini health-checkup stalls so people can get a quick checkup done to see if they test positive or negative for corona i.e., diagnosis is an Yes or No. Unfortunately, they could not have a doctor for each stall. So, they decided to have a decision tree where a person can be quickly tested based on the following attributes: (a) Country i.e., place where the person lives in, (b) Flu/Cough which indicates whether the person has Flu/Cough and (c) AS i.e., the person Appears Sick (AS).

Table below shows data provided by WHO. Let y1 be last digit of your roll number and y2 be the second last digit of your roll number. The number of data entries in the table is 10 with rows numbered from 1 to 10 (i.e., row 1 would be China, Yes, Yes, Yes). Please flip the value for Diagnosis in the rows identified using the following rules (i.e., make the Diagnosis value of Yes as No and vice versa). If two or more of the below rules identify the same row, please flip only once (e.g., if rules (i) and (ii) below identify same row, please flip only once). You need to the answer all questions using the modified table:

(i) 1 + [(y1 + y2)%10]

(ii) 1 + [(y2)%10]

(iii) 1 + [(y2+1)%10]

For example, if your roll number is 20191075, y1 would be 5 and y2 would be 7. Hence, diagnosis in row 3 would be No (flipped using rule (i)), row 8 would be Yes (flipped using rule (ii)) and row 9 would be Yes (flipped using rule (iii)).

Country	Flu/Cough	AS	Diagnosis
China	Yes	Yes	Yes
China	Yes	No	Yes
China	No	Yes	Yes
Italy	Yes	No	No
Italy	Yes	Yes	Yes
Italy	No	Yes	Yes
India	Yes	Yes	Yes
India	Yes	No	No
India	No	No	No
India	No	Yes	No

a) Please write down the values of y1 and y2 for your roll number. -- (2) points
b) What is the concept of Information Gain in decision trees ? How do you compute this ? -- (3) points
c) Please use the concept of Information Gain to identify which attribute (among Country, Flu/Cough, AS) would be picked as root node (i.e., which attribute would result in the highest gain at the root) ? Please show all the computations involved that led you to pick this attribute. -- (12) points