Indian Institute of Technology Kharagpur
Mid-Spring Semester 2019-20

Subject No.: CS60050 **Subject: Machine Learning** **Duration: 2 hours** **Full Marks: 50**
Department: Computer Science and Engineering
Specific charts, graph paper, log book etc., required: NA
Instructions:
This question paper contains two (02) printed pages.
Attempt all questions. All parts of the same question must be answered together.
No clarifications can be provided during the exam. Make reasonable assumptions if necessary, and state any assumptions made.
All workings must be shown. You can use calculators.

1. Consider Table 1 which consists of land prices based on the area and the proximity to city center. We want to build a linear regression estimator based on these data. We will use Mean Squared Error cost function. **[3 + 6 + 1 + 2 = 12]**

   (a) Write the hypothesis function, and the equations for updating the different parameters of this estimator.

   (b) Use the training data (first four rows of Table 1) to calculate the parameters after 2 iterations of gradient descent. Take the initial values of the parameters as 0.5 each, learning rate as 0.001. Show your steps.

   (c) Use the trained model to predict the price of the unknown data point (last row of Table 1) to the nearest million Rupees.

   (d) Mention two steps that can be taken to improve the prediction, apart from running Gradient Descent for more iterations.

Sr. No.	Area of Land (m^2)	Dist. to city center(Km)	Price (million INR)
1	25	12	13
2	30	10	18
3	21	25	9
4	28	2	17
5	35	15	?

Table 1: Land Prices

2. Answer the following questions. **[4 + 4 + 4 = 12]**

   (a) State two reasons why Linear Regression is not ideal for use in classification. Why is Mean Squared Error cost function not used with Logistic Regression? Write the cost function that is used instead.

   (b) Suppose you are given a model for binary (two-class) logistic regression. Explain briefly (e.g., 2-3 sentences each) how multi-class classification can be performed with the binary classifier using the following strategies:(i) One vs All, (ii) One vs One. If c is the number of classes, and m is the number of training examples given, mention how many instances of the binary classifier will be required in each of the above two cases.

(c) Consider Table 2 showing predictions by a classifier as well as the original labels. There are two classes – positive ('p') and negative ('n'). Calculate the following metrics with respect to class 'p': (i) Precision, (ii) Recall, (iii) True Positive Rate, and (iv) F1 score.

No.	Prediction	Actual Class
1	p	p
2	p	n
3	n	n
4	p	n
5	p	p

Table 2: Classifier Predictions and Actual Classes

3. (a) Derive a gradient descent training algorithm that minimizes the sum of squared error cost function, for the following hypothesis:
$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_1 x_1^3 + \theta_2 x_2 + \theta_2 x_2^3 + \ldots + \theta_n x_n + \theta_n x_n^3$
where $(x_1, x_2, \ldots, x_n)$ represents an instance having n features and $\theta_i, 0 \leq i \leq n$ represent the parameters to be learned. Assume that the number of training instances is m. Give your answer in the form $\theta_j \leftarrow \theta_j + \ldots$ for $1 \leq j \leq n$. **[6]**

(b) How can we increase the (1) Precision and (2) Recall for the positive class in (i) a Logistic Regression classifier, and (ii) a Decision Tree classifier? **[3+3=6]**

4. Consider Table 3 which contains data about second-hand car listings. We want to build a decision tree classifier to predict whether it is an acceptable deal to buy a listed second-hand car. We consider the two classes 'acc' and 'unacc'. ***[4 + 6 + 2 + 2 = 14]***

(a) Calculate the GINI index and the Entropy of the root node of a decision tree with the given training data (first seven rows of Table 3).

(b) Build the decision tree using the **ID3 algorithm using information gain**. Show all your steps.

(c) Use the decision tree to classify the unknown points in the table (last two rows of Table 3).

(d) Information Gain has a disadvantage that it prefers splits having large number of small but pure partitions. How do we overcome this?

No.	Pricing	Condition	Seating	Safety	Acceptibility
1	high	great	four	high	acc
2	low	good	four	low	acc
3	low	great	two	high	acc
4	low	good	two	high	unacc
5	high	bad	four	low	unacc
6	high	great	four	low	unacc
7	low	good	two	low	unacc
8	high	good	four	low	?
9	low	great	two	low	?

Table 3: Second hand car listings