Question Paper

Exam Date & Time: 06-May-2024 (02:30 PM - 05:30 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

VI SEMESTER B.TECH END SEMESTER EXAMINATIONS, APRIL-MAY 2024

MACHINE LEARNING [ICT 4032]

Α

Marks: 50

Answer all the questions.

Instructions to Candidates: Answer ALL questions Missing data may be suitably assumed

1)

2)

Consider the given dataset **Table Q1A** that represents the observation of the independent variable x versus the (5) dependent variable y:

A)

Table Q1A: Dataset

Х	Y
0.7913	-0.0085
-1.3752	-0.1450
-0.5895	0.2337
2.6625	0.1314
4.6816	1.4084
-0.8853	0.2992
1.4654	0.4463

- a. Using linear regression, find the equation of the best-fit line solution that represents the relationship between X and Y. Show your work and provide the equation of the regression line.
- b. Based on the equation obtained in part (a), predict the value of Y for X = 6. Explain your reasoning.
- B) Perform the K-means clustering for K=2, given the data {(0,0), (1,0), (0,1), (1,1), (1,2), (2,1), (2,2), (3,2), (6,6), (6 (3) ,7), (7,6), (7,7), (7,8), (8,6), (8,7), (8,8) (8,9) (9,7), (9,8), (9,9)}. Initial cluster centers are (8,8) and (9,9). Show the convergence for the above data. Justify your answer.
- C) List and briefly describe the different types of radial basis functions commonly used in interpolation and surface (2) fitting tasks.
- Minimization problems can be solved using the gradient descent method. Consider the **Table Q1A** and use the (5) gradient descent method to update the parameters W_0 and W_1 . Initially you may assume the value of $W_0=0.2$ A) and $W_1=0.1$. set the learning rate alpha as 0.1.
 - a. Iterate for 3 steps to update the parameters with suitable formula and calculations.
 - b. Why is the choice of learning rate of alpha being important in the gradient descent method? Discuss how different values of alpha affect the convergence of the algorithm.
- B) Given the Class Labeled Training Tuples from the AllElectronics Customer Database (Table 2B), analyze the (3) dataset to determine the information gain for potential splits on the attributes "age," "income," "student," and "credit rating" with respect to the class label "buys computer." Provide a step-by-step explanation of how you would calculate the information gain for each attribute and identify the attribute that yields the highest information gain.

Table 2B: Dataset

RID age

income

Duration: 180 mins.

	-0-				
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

How does the Machine Learning Paradigm differ from the Traditional Computing Paradigm? Illustrate the

(2)

Given a single layer perceptron with one output neuron X_1 receiving two inputs from input neurons X_2 and X_3 , (5) compute the update in weights for next three iterations with the following specification W_{12} =-3 W_{13} =2, X_1 =1, (5) X_2 =1, bias Θ =1, and desired output Y_1 =0.9. Assume sigmoid activation function and mean squared error as loss function.

- B) Assume we have a set of data from patients who have visited KMC hospital during the year 2021. A set of (3) features (e.g., temperature, height) have also been extracted for each patient. Our goal is to decide whether a new visiting patient has any of diabetes, heart disease, or Alzheimer (a patient can have one or more of these diseases). Some patient features are expensive to collect (e.g., brain scans) whereas others are not (e.g., temperature). Therefore, we have decided to first ask our classification algorithm to predict whether a patient has a disease, and if the classifier is 80% confident that the patient has a disease, then we will do additional examinations to collect additional patient features. In this case, which classification methods do you recommend: neural networks, decision trees, or naive Bayes?
- C) How does the utilization of ensemble methods, such as bagging and boosting, contribute to improving the (2) performance of machine learning models? Provide specific examples to illustrate the effectiveness of these techniques in real-world applications.
- 4) Consider the scenario where Bayesian decision theory is applied for pattern classification with continuous (5) feature spaces.

A) In this scenario, two classes, $\Omega = \{\omega 1, \omega 2\}$ are considered with class-conditional densities given by:

 $P(x | w1) \rightarrow N(2, 0.5)$ (Normal distribution)

distinctions using a flow diagram.

 $P(x \; | w2) \; \rightarrow N(1.5, \, 0.2)$

The prior probabilities are:

P(w1) = 2/3

C)

3)

P(w2) = 1/3

Additionally, the loss associated with each decision is represented by the matrix $\lambda = [1 2; 3 4]$.

Using the Bayesian formula and posterior probabilities, select the optimal decision with minimal risk to classify the input x=1.5.

- B) Illustrate the working of K-Nearest Neighbor classification. Mention any one advantage and one disadvantage (3) of this technique.
- C) What is the primary objective of Support Vector Machine (SVM) in classification tasks? Briefly explain the (2) concept of margin in SVM.
- 5) Consider the following Bayesian belief network representing causal relationship of five random variables, student's (5) intelligence (*I*), the course difficulty (*D*), the grade (*G*), the student's GATE score (*S*), and the quality of the recommendation
 A) letter (*L*). Given the conditional proability **Table Q5A**, find the proabability that a student gets a strong recommendation letter

Table Q5A



- B) Compare and contrast the parameter estimation and non-parameter estimation methods in statistics. Provide (3) an example of each method and explain how they are applied in practice.
- C) Describe the fundamental differences between supervised, unsupervised, and reinforcement learning in (2) machine learning. Provide a real-world scenario for each type and explain how the respective learning approach would be applied.

-----End-----