Question Paper

MANIPAL ACADEMY OF HIGHER EDUCATION

MANIPAL SCHOOL OF INFORMATION SCIENCES, MANIPAL
FIRST SEMESTER MASTER OF ENGINEERING - ME (BIG DATA AND DATA ANALYTICS / CLOUD COMPUTING / INTERNET OF THINGS / HEALTHCARE DATA ANALYTICS) DEGREE EXAMINATION - FEBRUARY 2021
Fundamentals of Machine Learning [BDA 601]

Marks: 100   **Duration: 180 mins.**

MONDAY, FEBRUARY 15, 2021

Answer all the questions.

1)
a.) Explain the difference between supervised and unsupervised machine learning? When should you use supervised over unsupervised machine learning? (L2) (TLO no 1.1) (5 marks) (5)

b.) Distinguish between classification and regression? When should you use classification over regression? (L2) (TLO no 1.1) (5 marks) (5)

2)
a.) Describe the method of choosing a Representation for the Target Function in designing a machine learning system. (L2)(TLO no 1.2)(5 marks) (5)

b.) Describe the method of choosing a Function Approximation Algorithm for a machine learning system. (L2)(TLO no 1.2)(5 marks) (5)

3) Explain the terms Concept Learning, Instance Space and Hypothesis Space. Also obtain Instance Space and Hypothesis Space for the following data. (2+2+2+2+2 Marks) (L2) (TLO no 2.1) (10)

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

- Instances X: Possible days, each described by the attributes
    - Sky (with possible values Sunny, Cloudy, and Rainy),
    - AirTemp (with values Warm and Cold),
    - Humidity (with values Normal and High),
    - Wind (with values Strong and Weak),
    - Water (with values Warm and Cool), and
    - Forecast (with values Same and Change).

4)
a.) Define entropy S of the Boolean classification and information gain of an attribute A, relative to a collection of examples S. (3 Marks) (L1) (TLO no 2.3) (3)

b.) Apply ID3 algorithm to construct a decision tree from the training examples given in the table (7 Marks) (L3) (TLO no 2.3) (7)

Instance	Classification	a_1	a_2

1	+	T	T
2	+	T	T
3	−	T	F
4	+	F	F
5	−	F	T
6	−	F	T

5)

a.) Explain about 'training Set' and 'test Set' in a Machine Learning Model? How much data will you allocate for the training, validation, and test Sets? (L2) (TLO no 2.3) (4 Marks) (4)

b.) Describe the K-Fold Cross-validation technique in training and testing with suitable example. (TLO no 2.3) (6 Marks)(L2) (6)

6) Solve for number of training examples "m" sufficient for successful learning of any target concept in the finite hypotheses space? (TLO no 3.1) (L3)(10 marks) (10)

7)

a.) Write the relationship between Bayes theorem and concept learning? (3 Marks)(L3) (TLO 4.1) (3)

b.) Produce the Brute-Force MAP LEARNING algorithm for concept learning to output the maximum a posteriori hypothesis (hMAP), based on Bayes theorem. (7 marks) (L3)(TLO 4.1) (7)

8) Write the purpose of confusion matrix and use the confusion matrix given below to calculate the following (L3) (TLO 4.3)(2+2+2+2+2 Marks) (10)

1. Accuracy
2. Misclassification Rate
3. Type-1 error and
4. Type-2 error

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

9) Solve for the principal components of the following covariance matrix (L3) (TLO 5.1)(10 Marks) (10)

$$\begin{pmatrix} 2 & -4 \\ -1 & -1 \end{pmatrix}$$

10) (4)

a.) Describe about "Ensemble models" with suitable example. (L2) (4 Marks) (TLO 5.2)

b.) Write the use of the following in ensemble models with suitable example (L3) (TLO 5.2) (6)

1. Averaging (2 Marks)
2. Majority vote and (2 Marks)
3. Weighted average (2 Marks)

-----End-----