## **Question Paper**

Exam Date & Time: 02-Jan-2023 (10:00 AM - 01:00 PM)



## MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences (MSIS), Manipal First Semester Master of Engineering - ME (Big Data Analytics) Degree Examination - January 2023

Fundamentals of Machine Learning [BDA 5103]

Marks: 100

Duration: 180 mins.

Monday, January 1, 2023

## Answer all the questions.

a) Write the principle of supervised and unsupervised learning <sup>(10)</sup> with a suitable example for each case. (6 marks) (L3) (TLO 1.1)

b) Write any two differences between machine learning and deep learning. (4 marks) (L3) (TLO 1.1)

- <sup>2)</sup> Write the description of the terms "Instance Space" and <sup>(4)</sup>
  <sup>a)</sup> "Hypothesis Space" in concept learning. (4 marks) (L3) (TLO 1.1)
  - <sup>b)</sup> Apply the FIND-S algorithm to obtain the hypothesis space for <sup>(6)</sup> the given training examples. (6 marks) (L3) (TLO 1.1)

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

3)

From the following training examples, compute entropy and <sup>(10)</sup> information gain values to select the root node of the decision tree (10 marks) (L3) (TLO 1.2)

Temperature	Outlook	Humidity	Windy	Play Golf?
hot	sunny	high	false	no
hot	sunny	high	true	no
hot	overcast	high	false	yes
cool	rain	normal	false	yes
cool	overcast	normal	true	yes
mild	sunny	high	false	no
cool	sunny	normal	false	yes
mild	rain	normal	false	yes
mild	sunny	normal	true	yes
mild	overcast	high	true	yes
hot	overcast	normal	false	yes
mild	rain	high	true	no
cool	rain	normal	true	no
mild	rain	high	false	yes

4)

Write the reasons for overfitting in decision tree models. How <sup>(4)</sup> will you address the overfitting issues in the decision tree (4 marks) (L3) (TLO 1.2)

<sup>b)</sup> Produce the solution to address the following issues in <sup>(6)</sup> decision tree learning:

a). Determining how deeply to grow the decision tree (3 marks) (L3) (TLO 1.2)b). Handling continuous-valued attribute (3 marks) (L3) (TLO 1.2)

- <sup>5)</sup> Relate overfitting and underfitting to bias and variance. Write <sup>(10)</sup> the role of hyper parameters in bias variance trade-off. (10 marks) (L3) (TLO 2.1)
- <sup>6)</sup> Compute the patient's probability of having liver disease if they <sup>(5)</sup> are an alcoholic based on the following data: The past data tells you that 10% of patients entering your clinic have liver disease, and five percent of the clinic's patients are alcoholics. Among those patients diagnosed with liver disease, 7% are alcoholics. If the patient is an alcoholic, what is the chance of having liver disease? (5 marks) (L3) (TLO 2.1)
  - <sup>b)</sup> Consider a hypothesis space containing three hypotheses: h1, <sup>(5)</sup> h2, and h3. Posterior probabilities of h1, h2, and h3 given the training data are 0.4, 0.3, and 0.3 respectively. When a new instance x is encountered, it is classified as by h1, but negative by h2 and h3. Apply Bayes' optimal classifier to obtain the most probable classification of the new instance. (5 marks) (L3) (TLO 2.1)
- <sup>7)</sup> Write the basic principle of case-based reasoning. Illustrate <sup>(10)</sup> the R4 model of the case-based reasoning (CBR) approach

with a suitable example. (10 marks) (L3) (TLO 2.2)

Expert assigned some basic outcomes to the nodes as given <sup>(10)</sup> in the table below for Bayesian Networks shown in the following figure. Apply joint probability to find the probability that it's raining when the yard is wet? (8 marks) (L3) (TLO 2.2)



9)

10)

Compute the Eigen values from the following data set: x = [0, (10) 8, 12, 20], y = [8, 9, 11, 12], and z = [10, 10, 10, 10]. (10 marks) (L3) (TLO 3.1)

- Relate the bagging and boosting ensemble models with a suitable example for each case. (4 marks) (L3) (TLO 3.2)
- <sup>b)</sup> Write the use of the following in ensemble models with suitable <sup>(6)</sup> example (L3) (TLO 3.2)
  - a) Averaging (2 marks)
  - b) Majority vote and (2 marks)
  - c) Weighted average (2 marks)

-----End-----