



ST JOSEPH'S UNIVERSITY, BENGALURU -27
M.Sc (BIG DATA ANALYTICS) – II SEMESTER
SEMESTER EXAMINATION: APRIL 2024

(Examination conducted in May / June 2024)

BDA 2321 – MACHINE LEARNING I
(For current batch students only)

Registration Number:

Date & session:

Time: 2 Hours

Max Marks: 50

This paper contains TWO printed pages and THREE parts

PART- A

Answer All questions

2X5=10

1. Define Machine learning with suitable example.
2. How Bayes theorem calculates posterior probability?
3. What is the entropy of a group in which:
 - a. All samples belong to the same class?
 - b. Each group having equal number of samples.
4. Why do we use cross-validation?
5. What is boosting?

PART - B

Answer any FIVE questions

4X5=20

6. Briefly explain supervised and unsupervised machine learning methods.
7. Differentiate PCA and ICA.
8. Explain the process of machine learning with example in each processing stages.
9. Differentiate between generative and discriminative learning models. In a multinational company, there are people speaking different languages of their own mother tongue. The auto teller engine hosted by the company has a task of determining the language that someone is speaking by determining the linguistic differences without learning any language. Which learning model it has to follow? Describe the model.
10. What is Hyperplane? Explain SVM model with an example
11. Explain any two Unsupervised learning methods in detail
12. Write a short note on LDA.

PART - C

Answer any TWO questions

10X2=20

13. NASA wants to be able to discriminate between Martians (M) and Humans (H) based on the following characteristics:

Green $\in \{N, Y\}$, Legs $\in \{2, 3\}$, Height $\in \{S, T\}$, Smelly $\in \{N, Y\}$.

The available training data is given in table below:

- a. Construct a decision tree using the ID3 algorithm.
- b. Write the learned concept for Martian as a set of conjunctive rules (e.g., if (green = Y and legs = 2 and height = T and smelly = N), then Martian; else Human).

	Species	Green	Legs	Height	Smelly
1	M	N	3	S	Y
2	M	Y	2	T	N
3	M	Y	3	T	N
4	M	N	2	S	Y
5	M	Y	3	T	N
6	H	N	2	T	Y
7	H	N	2	S	N
8	H	N	2	T	N
9	H	Y	2	S	N
10	H	N	2	T	Y

14. Differentiate overfitting and underfitting. How it can affect model generalization? Explain the different evaluation methods of classification.

15. Given 14 training examples of the target concert play tennis with attributes outlook, temperature, humidity and wind. The frequency of play tennis = 9 Frequency of not play tennis = 5 Conditional probabilities are given as

$$P(\text{outlook} = \text{rainy} | \text{Play} = \text{Yes}) = 2/9$$

$$P(\text{temp} = \text{cool} | \text{Play} = \text{Yes}) = 3/9$$

$$P(\text{humidity} = \text{High} | \text{Play} = \text{Yes}) = 3/9$$

$$P(\text{Windy} = \text{true} | \text{Play} = \text{Yes}) = 3/9$$

$$P(\text{Outlook} = \text{rainy} | \text{Play} = \text{No}) = 3/5$$

$$P(\text{temp} = \text{cool} | \text{Play} = \text{No}) = 1/5$$

$$P(\text{humidity} = \text{High} | \text{Play} = \text{No}) = 4/5$$

$$P(\text{Windy} = \text{true} | \text{Play} = \text{No}) = 3/5$$

Classify the new instance whether play = yes or No (Outlook = sunny, Temp = cool, Humidity = high, wind = strong) using Naive Bayes Classifier.