

IM.Sc DS Sem.-5 Examination

CC 304

Machine Learning-I

Time : 2-30 Hours]

November-2024

[Max. Marks : 70

Instructions: All questions are compulsory. Use of non-programmable scientific calculator is allowed.

- Q.1** (a) What is Machine Learning? Explain its types. (07)
 (b) Give the difference between AI & ML. (07)

OR

- (a) What is Supervised Machine Learning? Explain with Example. (07)
 (b) Explain any five applications of Machine Learning. (07)

- Q.2** (a) What is Regression? Explain in simple, multiple, and polynomial regression. (07)
 (b) What is Regularization? Explain types of regularization (07)

OR

- (a) Using Naïve Bayes estimate the conditional probabilities of each attribute (color, legs, height, smelly) for the species classes ('M', 'H') using the data given in the table. (07)
 Using these probabilities estimate the probability values for the new instance: (color = Green, legs = 2, Height = Tall, and smelly = No).
 Conclude new instance belongs to which species?

Sr. No.	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

- (b) Explain: K Nearest Neighbors (KNN) Algorithm (07)

- Q.3** (a) What is Over fitting and Under fitting? Explain with appropriate example. (07)
 (b) What is Gradient Descent? Explain in detail. (07)

OR

- (a) What is performance measurement? Explain four performance measurements. (07)
 (b) What is Machine Learning? How it differs from traditional programming? (07)

- Q.4** (a) What is Random Forest? Explain detail. (07)
 (b) Explain in detail: ID3 algorithm (07)

OR

(P.T.O.)

- (a) What is Bagging and Boosting? Explain with an example. (07)
- (b) Consider the following dataset for a binary classification task involving animals: (07)

Animal	Color	Size	Endangered
Lion	Yellow	Large	No
Elephant	Grey	Large	Yes
Rabbit	White	Small	No
Tiger	Orange	Large	Yes
Fox	Red	Medium	No
Panda	Black	Medium	Yes
Parrot	Green	Small	No
Giraffe	Yellow	Large	No

1. Calculate the Gini impurity for the entire dataset.
2. Evaluate the potential split based on the "Color" feature and calculate the Gini impurity for the resulting subsets, along with the overall Gini impurity after the split.

Q.5 Attempt any **SEVEN** out of **TWELVE**:

(14)

- (1) What is Scaling of data?
- (2) Explain: Pseudo labelling
- (3) Define: Confusion Matrix
- (4) What is R² or R squared?
- (5) Define: K in KNN
- (6) Explain the terminologies of SVM.
- (7) Explain: Elastic-Net Regression
- (8) What are the four assumptions of Linear Regression?
- (9) What is Random Forest?
- (10) Define: Boosting
- (11) What is CART Algorithm?
- (12) Define: Performance Measurement
