1206E462

Candidate's Seat No :_____

M.Sc. AIML & AIML (DS) Semester 2 Examination

Machine Learning

June-2024

Time : 3-00 Hours] **[Max. Marks : 100**

SECTION I

Q1.(a) Explain the following (ANY three): [09]

i. Feature Selection methods

ii. Outlier detection and handling

iii. Missing values handling

iv. Distance measures

Q1.(b) Explain k-means method of clustering . Write algorithm. How do u select value of k? [09]

Q2. With respect to SVM, answer the following: [16]

i. HyperPlane

ii. Support Vectors

iii. Kernel

iv. Margin

v. Gamma value

vi. C value

Q3 Why knn is called so? Explain knn method for classification with example. How do u find optimal value of k? does choosing a particular distance measure has any effect on performance of classifier? [16]

OR

Q3. What is ensemble approach? Give difference between Bagging and Boosting. Explain **ANY ONE** of given algorithms: [16]

i. AdaBoost

ii. Random Forest

(P.T.O)

SECTION II

Q4. Why Naïve Bayes algorithm is called so? For the given dataset, find probability of for each class for the instance X={Color=Green, Legs=2, Height=Tall, Smelly=No} [16]

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

Q5. For the given data set , find Entropy, information gain, Gini Index and Chi Sqaure values for the decision tree [16]

Outlook	Temperature	Humidity	Windy	Play?
sunny	hot	high	false	No
sunny	hot	high	true	No
overcast	hot	high	false	Yes
rain	mild	high	false	Yes
rain	cool	normal	false	Yes
rain	cool	normal	true	No
overcast	cool	normal	true	Yes
sunny	mild	high	false	No
sunny	cool	normal	false	Yes
rain	mild	normal	false	Yes
sunny	mild	normal	true	Yes
overcast	mild	high	true	Yes
overcast	hot	normal	false	Yes
rain	mild	high	true	No

Q6. Explain Simple Linear Regression. With example explain method of getting best fit line . Explain the evaluation parameters like R2, RSME and RSE [18]

OR

Q6. Explain the following: [18]

i. Bernaulli distribution and Bernaullis trials

ii. Logistic regression for classification