Total No. of Questions : 10]

SEAT No. :

P3176

[Total No. of Pages : 3

[5354]-693

B.E. (Information Technology)

MACHINE LEARNING

(2012 Pattern)

Time : 2½ Hours] *[Max. Marks :70*

Instructions to the candidates:

1) ***Neat diagrams must be drawn wherever necessary.***
2) ***Figures to the right side indicate full marks.***
3) ***Use of Calculator is allowed.***
4) ***Assume Suitable data if necessary.***

Q1) a) What is a cross-validation? How it improves the accuracy of the outcome? **[5]**

b) Define and explain Squared Error (SE) and Mean Squared Error (MSE) w.r.t. Regression. **[5]**

OR

Q2) a) What is a geometrical machine learning model? Explain it with one example. **[5]**

b) What is a polynomial regression? How it can be represented in a form of a matrix? **[5]**

Q3) a) Define following w.r.t. Binary Classification: **[4]**

i) True positive ii) False positive

iii) True Negative iv) False Negative

b) Suppose classifier's prediction is given as follows: **[6]**

		Predicted	
Actual		+	–
	+	60	15
	–	10	15

Calculate Accuracy, Weighted Accuracy, Precision and Recall for it.

OR

P.T.O.

Q4) a) What is a support vector ? How the margins are defined in support vector machine. **[4]**

b) What do you mean by zero centered and un-correlated features? What is the use of it in the solution of multivariate linear regression? **[6]**

Q5) a) Define Cluster Tree? Write and explain Agglomerative Clustering algorithm. **[9]**

b) A survey is conducted on two attributes 'Acid Durability (secs)' and 'Strengths(kg/m^2)' to classify whether a special paper tissue is 'good' or not. Following are the samples received. **[9]**

Sr. No.	Acid Durability (secs)	Strength (kg/m^2)	Target Class
1	7	7	Bad
2	7	4	Bad
3	3	4	Good
4	1	4	Good
5	2	5	Good

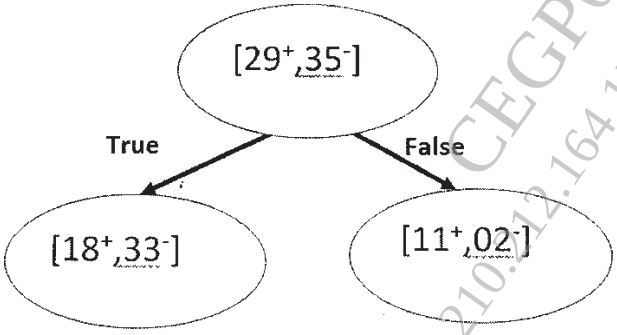
If a particular sample is surveyed with 'Acid Durability' = 3 seconds and 'Strength=7 kg/m^2', then what will be its target class if value of k=2 and k=3?

OR

Q6) a) Define with respect to Association Rule Mining: **[8]**

i) Support
ii) Confidence
iii) Lift

b) Consider a dataset of 29 positive samples and 35 negative samples. The dataset is split on certain condition. The split is as shown in the figure. Calculate "Information Gain" of the given split. **[10]**



Q7) a) Define and explain: **[8]**

i) Burnoulli's Distribution

ii) Binomial Distribution

b) Write and explain Naïve Bayes Classification Algorithm. **[8]**

OR

Q8) a) Describe Normal Distribution with its features. **[8]**

b) Approximately 1% of population among the age group of 40-50 have diabetis. A person with diabetis has a 90% of chance of a positive test, while a person without diabetis have 9% of chance of false positive result. What is the probability that a person has diabetis given that she just had a positive test? **[8]**

Q9) a) Explain Deep Learning. What are the challenges in Deep Learning? **[7]**

b) Write a short note on: **[9]**

i) Sequence Prediction

ii) Sequence Classification

iii) Sequence Generation

OR

Q10)a) Can we ensemble multiple models of same Machine Learning algorithm? Describe. **[8]**

b) Explain Multi-task learning with Task grouping and overlap methodology. **[8]**

❖ ❖ ❖ ❖