Total No. of Questions: 6 *Total No. of Printed Pages:3*

Enrollment No.....

Faculty of Science / Engineering

End Sem Examination May-2024

BC3CO59 / OE00096 Advanced Machine Learning

Programme: B.Sc. Branch/Specialisation: Computer Science / All

Duration: 3 Hrs. **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

Q.1 i. Which of the following is a common method for splitting nodes in a decision tree? **1**

(a) Gini impurity (b) Cross validation

(c) Gradient descent (d) Principal component analysis

ii. What is the purpose of pruning in decision trees? **1**

(a) To reduce the depth of the tree and prevent overfitting.

(b) To optimize the tree's parameters.

(c) To handle missing data.

(d) To improve the tree's interpretability

iii. What is the primary difference between bagging and boosting in the context of decision trees? **1**

(a) Bagging trains multiple trees independently, while boosting trains trees sequentially

(b) Bagging improves interpretability, while boosting improves predictive accuracy.

(c) Bagging reduces computational complexity, while boosting increases it.

(d) Bagging handles missing data, while boosting does not

iv. If a model predicts a negative class for a data point that belongs to the positive class, what term is used in the confusion matrix? **1**

(a) True positive (b) False negative

(c) True negative (d) Misclassification

v. What type of variable is the target variable in binary logistic regression? **1**
(a) Continuous
(b) Binary
(c) Categorical (with more than two classes)
(d) Ordinal

vi. What does the odds ratio tell us in logistic regression? **1**
(a) The absolute difference in probability for a unit change in the predictor.
(b) The direction (positive or negative) of the relationship between the predictor and the outcome.
(c) The factor by which the odds of the event occurring are multiplied for a unit change in the predictor.
(d) The p-value associated with the significance of the predictor.

vii. What is the range of possible values for AUC? **1**
(a) 0% to 50%
(b) -1 to 1
(c) 0 to 1
(d) There is no upper limit, but the minimum is 0.

viii. What is the relationship between AUC and the ROC curve? **1**
(a) They are completely unrelated concepts.
(b) AUC is a specific point on the ROC curve.
(c) AUC represents the total area under the ROC curve.
(d) A higher AUC always corresponds to a steeper ROC curve.

ix. The effectiveness of an SVM depends upon _____. **1**
(a) Selection of Kernel trick (b) Kernel parameters
(c) Soft margin parameter C (d) All of these

x. What is the name of the line that separates different classes in SVM? **1**
(a) Decision boundary (b) Separator line
(c) Class divider (d) Margin line

Q.2 i. What do you mean by classification tree? **2**

ii. Differentiate between training and testing data set. **3**

iii. What are the major pruning techniques used in decision tree? Explain with suitable example. **5**

OR iv. Compare the different attribute selection measures in decision tree. **5**

Q.3 i. Discuss the significance of F measure in machine learning. **2**

ii. What is K fold cross validation test in machine learning? Explain with suitable example. **8**

OR iii. Explain confusion matrix with suitable example. **8**

Q.4 i. What do you mean by logistic regression? Explain with suitable example. **3**

ii. Compare and contrast linear regression and logistic regression. Explain different types of logistic regression. **7**

OR iii. What is probit and logit analysis? Explain. **7**

Q.5 i. What is area under the curve (AUC) and why it is used? **4**

ii. Which is the likelihood ratio test for goodness of fit? Explain with suitable example. **6**

OR iii. Elaborate Pseudo R square tests and why it is used? Explain. **6**

Q.6 Attempt any two:

i. What are some common kernel functions used in SVMs? **5**

ii. Define kernel trick in SVM. **5**

iii. Discuss hard margin and soft margin SVM. **5**

******

Marking Scheme
Advanced machine learning-BC3CO59(T)

Q.1	i)	(a) Gini impurity	**1**
	ii)	(a) To reduce the depth of the tree and prevent overfitting.	**1**
	iii)	(a) Bagging trains multiple trees independently, while boosting trains trees sequentially	**1**
	iv)	(b) False negative	**1**
	v)	(b) Binary	**1**
	vi)	(c)The factor by which the odds of the event occurring are multiplied for a unit change in the predictor.	**1**
	vii)	(c) 0 to 1	**1**
	viii)	(c) AUC represents the total area under the ROC curve.	**1**
	ix)	(d) All of the above	**1**
	x)	(a) Decision Boundary	**1**
Q.2	i.	Define entropy in the context of decision trees?	**2**
	ii.	List down the different types of nodes in Decision Trees.	**3**
	iii.	What are the major pruning techniques used in decision tree? Explain with suitable example.	**5**
		Types- 3 Marks	
		Example – 2 Marks	
OR	iv.	Compare the different attribute selection measures in decision tree.	**5**
		1 measure – 2 Marks	
Q.3	i.	What is the significance of F measure in Machine learning?	**2**
	ii.	What is K fold cross validation test in Machine learning? Explain with suitable example.	**8**
		K fold – 5 Marks	
		Example – 3 Marks	
OR	iii.	Define Confusion matrix with suitable example.	**8**
		cm- 5 Marks	
		example- 3 Marks	
Q.4	i.	What do you mean by logistic regression? Explain with suitable example.	**3**
		Logistic regression – 2 Marks	
		Example – 1 Marks	
	ii.	Compare and contrast linear regression and logistic regression. Explain different types of logistic regression.	**7**
		Compare and contrast – 3 Marks (1 for each)	
		Type – 4 Marks (1 for each type)	
OR	iii.	What is probit and logit analysis? Explain.	**7**
		Probit – 4 Marks	
		Logit – 3 Marks	
Q.5	i.	What is area under the curve (AUC) and why it is used?	**4**
		What – 2 Marks	
		Why – 2 Marks	
	ii.	What is the likelihood ratio test for goodness of fit? Explain with suitable example.	**6**
		Test – 4 Marks	
		Example – 2 Marks	
OR	iii.	What is Pseudo R square tests and why it is used? Explain.	**6**
		Test – definition – 3 Marks	
		Algorithm – 3 Marks	
Q.6		Attempt any two:	
	i.	What are some common kernel functions used in SVMs?	**5**
	ii.	Define kernel trick in SVM?	**5**
	iii.	Define hard margin and soft margin SVM?	**5**
		Hard margin – 2.5 Marks	
		Soft margin – 2.5 Marks	

******