

2024

COMPUTER SCIENCE AND ENGINEERING

Paper : CSE-901

[Machine Learning (Elective-I)]

Full Marks : 70

*The figures in the margin indicate full marks.**Candidates are required to give their answers in their own words as far as practicable.*Answer **question nos. 1 & 2**, and **any four** questions from the rest.1. Answer **any five** questions :

2×5

- What is the necessity of dimensionality reduction in machine learning?
- What are Loss Function and Cost Functions? Explain the key difference between them.
- How do you select essential variables while working on a data set?
- How are covariance and correlation different from one another?
- When does regularization, normalization, or standardization come into play in Machine Learning?
- A data set about utility fraud detection has been given to you. You have built a classifier model and achieved a performance score of 96.5%. Is this a good model? If yes, justify. If not, what can you do about it?
- Explain the phrase 'Curse of Dimensionality'.

2. Answer **any five** questions :

4×5

- Define 'decision tree'. Construct the decision tree for the following Boolean functions:
 - $A \text{ XOR } B$
 - $A \cap [B \cup C]$.
- Suppose the classifier's prediction is given as follows :

		Predicted	
		+	-
Actual	+	60	15
	-	10	15

Calculate accuracy, weighted accuracy, precision, and recall.

- What is cross-validation? How does it improve the accuracy of the outcomes?

Please Turn Over
(8732)

- (d) How do you justify the convergence criteria of any cost function for regressions?
- (e) State two reasons why linear regression is not ideal for classification. Why is the Mean Squared Error cost function not used with Logistic Regression? Write the cost function that is used instead.
- (f) What is a Random Forest? How does it work?
- (g) What are the different terminating conditions for decision trees? Explain your answer.
3. (a) Build and explain the decision tree upto 2 levels for the following transactions :

Tid	Refund	Material Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Married	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- (b) What is the information gain of A_1 and A_2 relative to the transactions in the table?

Instance	1	2	3	4	5	6	7	8	9
A_1	T	T	F	F	F	F	F	T	F
A_2	T	T	F	F	T	T	F	F	T
Target Class	+	+	-	+	-	-	-	+	-

6+4

4. (a) Explain the assumptions of linear regression. Why are they important?
- (b) Derive the formula for the least squares method used in linear regression.
- (c) Discuss the potential issues of multicollinearity in linear regression. How can it be detected and resolved?
- (d) What are the common evaluation metrics used for linear regression? Explain each with formulas.
- 2+3+3+2
5. (a) Compare linear regression with logistic regression. When should each be used?
- (b) Derive the likelihood function and the cost function for logistic regression.
- (c) Explain the methods to evaluate the performance of a logistic regression model.
- 3+3+4

6. (a) What is a K-nearest neighbor classifier?
(b) Write down the KNN algorithm pseudocode and comment on the algorithm's performance on different data volumes.
(c) What are the advantages and limitations of this algorithm?
(d) What are the various types of KNN algorithms? 2+4+2+2
7. (a) What is meant by feature extraction? Write an algorithm to extract the features from a given dataset.
(b) Critically comment on a recommender system. Briefly explain the different phases of a recommendation system. 5+5
8. (a) How can you select K for K-means Clustering?
(b) What are support vectors in SVM, and what are different kernels in SVM?
(c) What are Loss functions and Cost Functions with respect to SVM? Explain the critical difference between them. 2+4+4
9. (a) Discuss how the feedforward process works in a neural network.
(b) Explain the backpropagation algorithm in detail. How does it update weights in a neural network?
(c) What are different activation functions, and what are their significances? 3+4+3
-