B.Tech Even Semester Examination, May, 2025

Computer Science & Engineering
(6th Semester)

Course No: CSE-603(21)
(Machine Learning)

Full Marks: 70
Pass Marks: 28
Time: 3 hours

Note:
1. Attempt any five questions taking one from each unit.
2. Begin each answer in a new page
3. Answer parts of a question at a place
4. Assume reasonable data wherever required
5. The figure in the right margin indicates full marks for the question
6. All the mathematical symbols and abbreviations have their usual meanings.

UNIT-I

1. a) What's the trade-off between bias and variance? 3

   b) What's the difference between a generative and discriminative model? 2

   c) When should you use classification over regression? 2

   d) What is the difference between covariance and correlation? 3

   e) What are the basic assumptions to be made for linear regression? 4

2. a) What is underfitting andoverfitting in Machine Learning and how do we prevent it? 4

Turn Over

b) How is a decision tree pruned? 4

c) Given the dataset below to learn a decision tree which predicts if people pass (Yes or No), based on their previous GPA and whether or not they studied.

GPA	Studied	Passed
L	F	F
L	T	T
M	F	F
M	T	T
H	F	T
H	T	T

What is the entropy H(Passed)?

What is the entropy H(Passed | GPA)?

What is the entropy H(Passed | Studied)? 6

UNIT-II

3. a) Why is it called Logistic Regression if it is a classification model? What are the outputs of a logistic regression model? 3

b) Explain the Gradient Descent algorithm. Write down the update formula in each iteration. Explain the effects of high and low learning rates. 4+2

c) Below is a diagram if a single artificial neuron (unit):

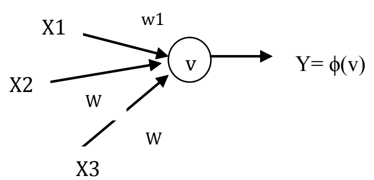

Figure 1: Single unit with three inputs.

UNIT-V

9. a) Define the components of a Markov Decision Process (MDP). How is it used in reinforcement learning? 7

b) Differentiate between value-based and policy-based reinforcement learning methods. 4

c) What is the Bellman Equation? Write the Bellman Expectation Equation for the value function. 3

10. a) Explain, in words, the Exploration-exploitation tradeoff in Reinforcement Learning. Provide one example of Exploration-Exploitation in the real world. 5

b) What is Q-Learning? How Does Q-Learning Work? 9

*****

The node has three inputs x = (x1, x2, x3)1z that receive only binary signals (either 0 or 1). How many different input patterns this node can receive? What if the node had four inputs? Five? Can you give a formula that computes the number of binary input patterns for a given number of inputs? 3

d) Suppose we have a multilayer perceptron (MLP) model with 17 neurons in the input layer, 25 neurons in the hidden layer and 10 neuron in the output layer. What is the size of the weight matrix between the hidden layer and output layer? 2

4. a) What are missing values? How do you handle missing values? 5

b) What is the Naive assumption in the Naive Bayes Models? At what stage will you use Bayes rule in this algorithm? 3

c) What is advantage of using neural networks over linear regression? 2

d) What is a kernel in SVM? Why do we use kernels in SVM?UNIT III 4

5. a) What is the general principle of an ensemble method and what is bagging and boosting in ensemble method? 6

b) Explain the difference between a test set and a validation set? 3

c) Explain the forward feature selection and backword feature selection methods. 5

6. a) Why is feature extraction important? 3

b) Why is model evaluation important in Machine Learning? 2

c) You have trained a supervised learning model for a binary classification task. The model shows high accuracy on the training data but low accuracy on the test data.

(a) Explain two possible causes for this discrepancy between training and test performance.

(b) Describe three techniques you would use to evaluate and debug the performance of your learning algorithm. 9

UNIT-IV

7. a) Write the Expectation-Maximization (EM) algorithm 5

b) Write 3 similarities and 3 differences between Kmeans algorithm and EM algorithm 6

c) Explain the Local Minima issue in the K-Means clustering algorithm 3

8. a) How do you choose the number of clusters in clustering algorithms 3

b) What is a dendrogram and how is it used in hierarchical clustering? 3

c) What does it mean for the principal components to be “orthogonal,” and why is this property important in PCA? 3

You apply PCA to a dataset with 200 features and find that the first 20 components explain 98% of the variance.

(a) What does this tell you about the data?

(b) How might this result influence your approach to training a machine learning model? 5