Department of Computer Science & Engineering
Quiz-I, Monsoon Semester, 2025-26
Subject: Machine Learning (NCSD 519)

Time: 30 Minutes **Full Marks: 20**

Name of Student: Admission No.:

Instruction: (i) Q.No.1 to 4 carries 1 Marks
(ii)Q. No. 5 to 12 carries 2 Marks
(iii) More than one option may be correct.

---

1. Regarding bias and variance, which of the following statements is true?
(a) Models which overfit have a high bias and underfit have a high variance
(b) Models which overfit have a high bias and underfit have a low variance
(c) Models which overfit have a low bias and underfit have a high variance
(d) Models which overfit have a low bias and underfit have a low variance

2. Which loss function is used in logistic regression?

(a) Mean Squared Error (MSE) (b) Cross-Entropy
(c) Maximum Likelihood Estimation (MLE) (d) Absolute Error

3. Which of the following best describes the role of unlabelled data in semi-supervised learning?
(a) It lowers the model's predictive performance
(b) It increases the cost of data collection
(c) It helps capture hidden patterns and improves generalization
(d) It completely removes the need for labelled data

4. A decision tree model is trained until it classifies every training sample correctly, without any restriction on depth. What is the most likely outcome of deep depth?
(a) The model will underfit the data (high bias, low variance)
(b) The model will overfit the data (low bias, high variance)
(c) The model will generalize well (low bias, low variance)
(d) The model will fail to learn patterns (high bias, high variance)

5. . A company trains a Support Vector Machine (SVM) to classify emails as spam or not spam. Since some emails are ambiguous, the dataset is not perfectly separable. The engineers introduce slack variables into the model. Which of the following statements are correct? (Select all that apply)

(a) Slack variables allow some misclassifications to achieve a softer margin
(b) With slack variables, the optimization balances maximizing margin and minimizing classification errors
(c) If the cost parameter C is set very high, the model will tolerate more slack(misclassification)
(d) If the cost parameter C is small, the model allows more slack but gains better generalization

6. A dataset has 12 samples (8 Positive, 4 Negative). What is the **entropy** of this dataset?

(a) 0.918
(b) 0.811
(c) 1.000
(d) 0.722

7. If predicted probability p=0.8 and true label y=1, cross-entropy loss is:
(a) 0.22
(b) 0.10
(c) 0.50
(d) 1.00

8. What is the margin in SVM? If w=(3,4)

(a) 0.2
(b) 0.4
(c) 0.5
(d) 0.6

9. A company is developing an autonomous car to navigate city streets safely. The car has:

- 500 hours of labelled driving data showing correct actions in certain situations,
- 9,500 hours of unlabelled driving video data,
- A feedback system that gives rewards for safe driving (+10) and penalties for collisions (-5).

Which ML approaches could be suitable for training the car?

(a) Supervised learning
(b) Semi-supervised learning
(c) Reinforcement learning
(d) Unsupervised learning

10. You train an SVM on a gene expression dataset with 20,000 features. Despite the high dimensionality, the model performs well. The main reason is:
(a) SVM reduces data to the top 2 principal components
(b) SVM's decision boundary relies only on a subset of training points
(c) SVM always applies dimensionality reduction
(d) SVM maximizes the margin, which helps generalization in high dimensions

11. Dataset: (1,3),(2,6)

Initialize w=0,b=0. Learning rate η=0.2.
Perform 1 epoch batch gradient descent. What is the updated slope w?
(a) 1.2
(b) 1.8
(c) 2.4
(d) 3.0

12. Logistic regression model outputs a probability of 0.85 for a patient having cancer. Which interpretations are valid? (Select all that apply)

(a) The patient definitely has cancer
(b) The patient has an 85% estimated chance of cancer according to the model
(c) If threshold = 0.5, the model will classify the patient as “cancer present”
(d) This probability is based on the learned coefficients and input features

Answer Sheet

Q. No.	Write Correct Option
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	