Machine Learning Supervised Learning Quiz - Set 1

Instructions:

- Answer all questions clearly and completely.
- Show your work for subjective questions.
- For multiple choice questions, circle the correct option.
- **Total marks: 79** (MCQ: 21 marks, Subjective: 58 marks)

Multiple Choice Questions

Q1. Which regularization technique adds a penalty term proportional to the sum of absolute values of parameters? **[2 marks]**

(A) L1 Regularization (Lasso)

(B) Dropout

(C) Elastic Net

(D) L2 Regularization (Ridge)

Q2. In Support Vector Machines, what happens when the regularization parameter C is very large? **[3 marks]**

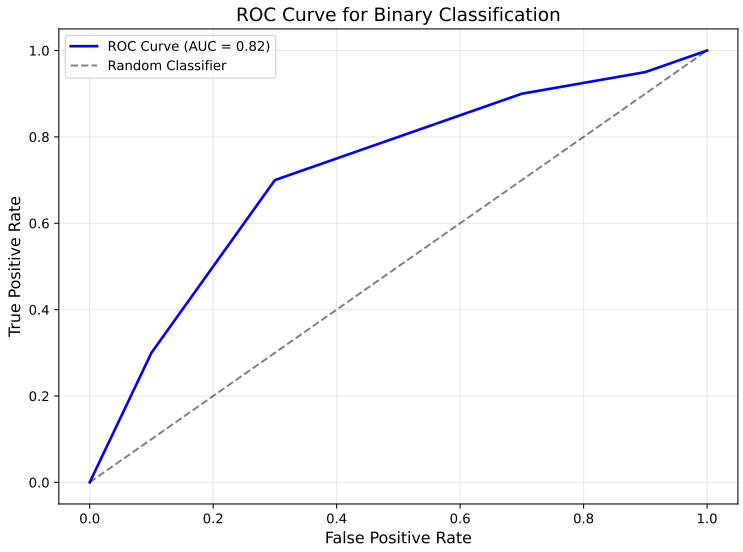
(A) The model focuses on minimizing training error and may overfit

(B) The kernel function becomes linear

(C) The model becomes more regularized and may underfit

(D) The support vectors are ignored

Q3. Looking at the ROC curve below:



What does the area under the curve (AUC) represent? **[3 marks]**

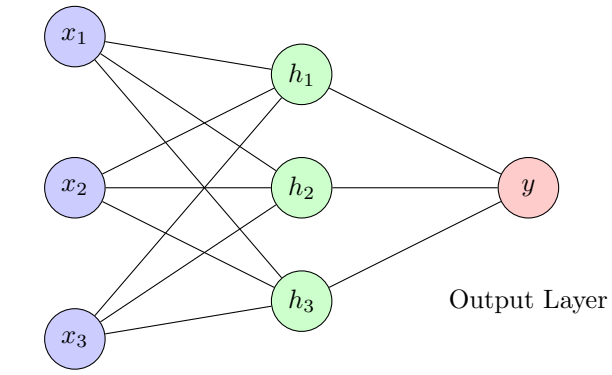
(A) The total number of correct predictions

(B) The difference between true positive rate and false positive rate

(C) The computational complexity of the algorithm

(D) The probability that the classifier ranks a random positive instance higher than a random negative instance

Q4. In the neural network architecture shown:



How many parameters (weights and biases) does this network have? **[4 marks]**

(A) 13

(B) 16

(C) 15

(D) 12

Q5. In a decision tree, which impurity measure is most commonly used for classification tasks? **[2 marks]**

(A) Mean Absolute Error (MAE)

(B) Gini Impurity

(C) Mean Squared Error (MSE)

(D) R-squared

Q6. Which of the following is true about k-fold cross-validation? **[2 marks]**

(A) It splits data into k parts, trains on k-1 parts, tests on 1 part, repeats k times

(B) It requires k different algorithms

(C) It uses k different datasets for training

(D) It only works when k equals the number of features

Q7. Which of the following best describes the bias-variance tradeoff in machine learning? **[2 marks]**

(A) High bias models always perform better than high variance models

(B) Reducing bias typically increases variance, and vice versa

(C) Bias and variance are independent and don't affect each other

(D) Variance only matters in unsupervised learning

Q8. Given the confusion matrix below for a binary classification problem:

	Predicted 0	**Predicted 1**
Actual 0	85	15
Actual 1	10	90

What is the precision of the classifier? **[3 marks]**

(A) 0.90

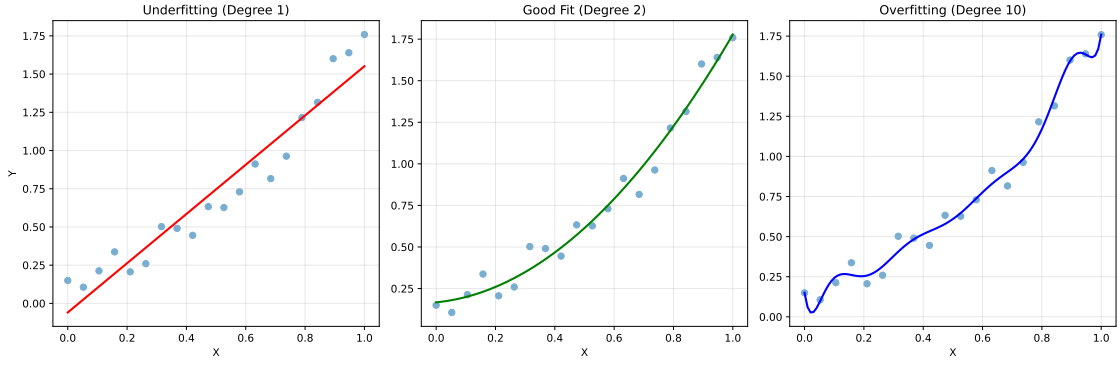
(B) 0.875

(C) 0.85

(D) 0.857

Subjective Questions

Q9. **[Total: 7 marks]** Examine the overfitting comparison plots:



a) Identify which model suffers from underfitting, good fit, and overfitting. Justify each choice. **[4 marks]**

b) If you had to choose a regularization parameter λ for ridge regression, would you choose a high or low value to move from the overfitted model to the good fit model? Explain. **[3 marks]**

Q10. **[Total: 9 marks]** Compare and contrast the following three supervised learning algorithms in terms of their assumptions, strengths, and weaknesses:

Algorithm	**Key Assumptions**	**Strengths**	**Weaknesses**
Linear Regression			
Decision Trees			
k-NN			

Fill in the table above and provide a brief explanation for each entry. **[9 marks]**

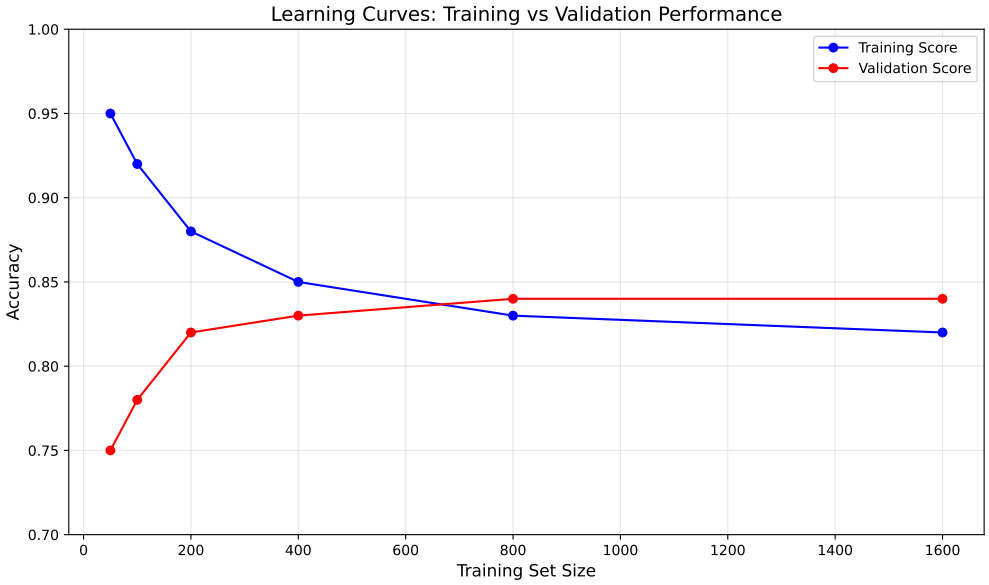
Q11. [Total: 10 marks] Explain the ensemble methods Random Forest and AdaBoost:

a) Describe how Random Forest reduces overfitting compared to a single decision tree. **[3 marks]**

b) Explain the boosting principle in AdaBoost and how it differs from bagging. **[4 marks]**

c) Under what circumstances would you prefer Random Forest over AdaBoost and vice versa? **[3 marks]**

Q12. [Total: 6 marks] Given the learning curves shown below:



a) Identify whether the model is suffering from high bias or high variance. Justify your answer. **[3 marks]**

b) Suggest two specific techniques to improve the model performance. **[2 marks]**

c) If the training accuracy is 82% and validation accuracy is 79%, calculate the overfitting gap. **[1 mark]**

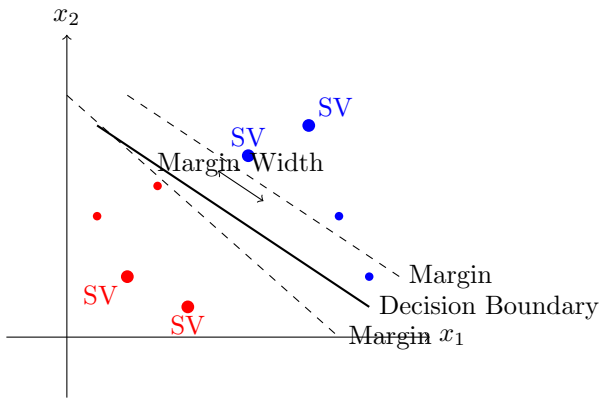
Q13. [Total: 8 marks] Derive the gradient descent update rule for logistic regression. Start from the logistic loss function:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))]$$

where $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$

Show all steps clearly including the partial derivatives. **[8 marks]**

Q14. [Total: 7 marks] Consider the SVM with margin illustration:



a) Explain the concept of support vectors and their role in SVM. **[3 marks]**

b) If we have 120 positive samples and 80 negative samples, and 15 of them are support vectors, what percentage of the data points are support vectors? **[2 marks]**

c) Compare the computational complexity of SVM prediction with and without kernel tricks. **[2 marks]**

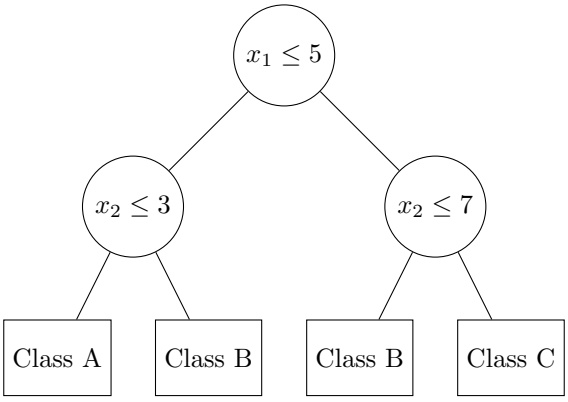
Q15. **[Total: 5 marks]** Consider the following dataset for linear regression:

Sample	**Feature 1**	**Feature 2**	**Target**
1	1	2	6
2	3	1	7
3	2	3	9
4	4	2	10

a) Calculate the mean squared error (MSE) if the model predicts $\hat{y} = 5.9, 7.1, 8.8, 9.9$ respectively. **[3 marks]**

b) If we use L2 regularization with $\lambda = 0.05$, write the complete loss function. **[2 marks]**

Q16. **[Total: 6 marks]** Analyze the decision tree structure below:



a) What is the maximum depth of this tree? **[1 mark]**

b) Calculate the Gini impurity for a node with class distribution: Class A: 40 samples, Class B: 30 samples, Class C: 10 samples. **[3 marks]**

c) Explain why pruning might be beneficial for this tree. **[2 marks]**