SL TEST 2 (20 to 30 mins)



15-Question Test on Overfitting & Underfitting,   
Regularization, Ridge Regression, Lasso Regression, and Logistic Regression



Section 1: Overfitting & Underfitting

Q1. Define overfitting and underfitting in the context of machine learning.

Answer:  **Overfitting** occurs when a model learns the noise and details of the training data too well, resulting in poor generalization to new data.

 **Underfitting** happens when a model is too simple to capture the underlying patterns, leading to poor performance on both training and test data.

Q2. What are the signs that a model is overfitting?   
Answer: High accuracy on training data but poor accuracy on test/validation data

* Large gap between training and validation performance
* Complex models with many parameters relative to the data size

Q3. What are some common methods to prevent overfitting? Answer:  Use regularization (L1/L2)

 Reduce model complexity

 Increase training data

 Use cross-validation

 Apply dropout (in neural networks)

 Early stopping during training

Section 2: Regularization

Q4. What is regularization, and why is it used in machine learning models? Answer: Regularization is a technique used to prevent overfitting by adding a penalty to large coefficients in the model. It discourages complex models and helps improve generalization.

Q5. Explain the difference between L1 and L2 regularization.

Answer:  **L1 Regularization (Lasso)** adds the absolute value of coefficients as a penalty term. It can shrink some coefficients to zero, performing feature selection.

 **L2 Regularization (Ridge)** adds the squared value of coefficients. It shrinks coefficients smoothly but doesn’t make them exactly zero.

Q6. How does regularization affect the bias-variance tradeoff? Answer: Regularization increases **bias** slightly by simplifying the model but reduces **variance** significantly. This helps achieve better generalization by reducing overfitting.

Section 3: Ridge Regression

Q7. What is Ridge Regression, and how does it differ from standard linear regression? Answer: Ridge Regression is a linear regression technique that includes an L2 penalty term. It differs from standard linear regression by penalizing large coefficients, which helps reduce overfitting.

Q8. What is the effect of the regularization parameter (λ) in Ridge Regression?

Answer: The parameter λ\lambdaλ controls the strength of the penalty.

* **Large λ:** More regularization → smaller coefficients
* **Small λ:** Less regularization → behaves like ordinary linear regression  
  Proper tuning of λ balances bias and variance.

Section 4: Lasso Regression

Q9. What is Lasso Regression, and how does it perform feature selection?

Answer: Lasso Regression uses L1 regularization, which can shrink some coefficients to exactly zero. This effectively removes less important features, hence performing feature selection automatically.

Q10. In which scenarios is Lasso Regression preferred over Ridge Regression?

Answer: Lasso is preferred when:

* Feature selection is desired
* Some input features are irrelevant
* The dataset has more features than observations  
  Ridge is better when all features are useful but correlated.

Section 5: Logistic Regression

Q11. Explain the purpose of Logistic Regression.

Answer: Logistic Regression is used for binary (or multi-class) classification tasks. It models the probability that a given input belongs to a particular class using the sigmoid function.

Q12. How does the cost function in Logistic Regression differ from that in Linear Regression?

Answer:  **Linear Regression** uses Mean Squared Error (MSE) as its cost function.

 **Logistic Regression** uses **log loss (cross-entropy loss)**, which is better suited for probability-based classification.

Section 6: Multiple Choice Questions

Q13. Which of the following statements about Lasso Regression is true?

A) Lasso Regression always includes all features in the final model.

B) Lasso Regression can set some coefficients to exactly zero, performing feature selection. C) Lasso Regression is not affected by the choice of the regularization parameter.

D) Lasso Regression can only be applied to linear models.

Answer B) Lasso Regression can set some coefficients to exactly zero, performing feature selection.

Q14. What is the main disadvantage of using Ridge Regression compared to Lasso Regression?

A) It cannot handle multicollinearity.

B) It does not perform feature selection.

C) It requires more computational resources.   
D) It can only be used for binary classification.

**Answer:**  
**B) It does not perform feature selection.**

Q15. Explain the concept of the confusion matrix and its significance in evaluating the performance of classification models like Logistic Regression.

Answer:

A **confusion matrix** is a table used to evaluate the performance of a classification model by showing the counts of:

* **True Positives (TP)**
* **False Positives (FP)**
* **True Negatives (TN)**
* **False Negatives (FN)**

It helps compute metrics like accuracy, precision, recall, and F1-score, which give a more detailed understanding of model performance than accuracy alone.