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 training data too well, including its noise and outliers, leading to poor generalization on unseen data.

**Underfitting** happens when a model is too simple to capture the underlying patterns in the data, resulting in poor performance on both training and test sets

Q2. What are the signs that a model is overfitting?   
Answer:

High accuracy on the training set but low accuracy on the validation/test set.

A large gap between training and validation loss.

Model performs well on seen data but poorly on new, unseen data.

Q3. What are some common methods to prevent overfitting? Answer:

Use regularization (L1 or L2).

Employ cross-validation.

Simplify the model (fewer parameters or layers).

Add more training data or use data augmentation.

Use dropout (for neural networks).

Apply 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 reduce overfitting by adding a penalty term to the loss function, discouraging overly complex models by shrinking the model coefficients.

Q5. Explain the difference between L1 and L2 regularization. Answer:

**L1 Regularization (Lasso):** Adds the absolute value of the coefficients to the loss function. Can shrink some coefficients to zero, effectively performing feature selection.

**L2 Regularization (Ridge):** Adds the squared value of the coefficients. Shrinks coefficients but does not eliminate them.

Q6. How does regularization affect the bias-variance tradeoff? Answer:

Regularization increases bias slightly but reduces variance significantly, helping to prevent overfitting and improve generalization.

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 model that includes an L2 regularization term to penalize large coefficients. It differs from standard linear regression by preventing overfitting and handling multicollinearity better.

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

The regularization parameter λ\lambdaλ:

* Controls the strength of the penalty.
* A higher λ\lambdaλ shrinks coefficients more, increasing bias but reducing variance.
* A lower λ\lambdaλ makes the model closer to standard linear regression.

Section 4: Lasso Regression

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

Answer:

Lasso Regression is a linear model with L1 regularization. It can shrink some coefficients to exactly zero, effectively removing irrelevant features and thus performing feature selection.

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

Answer:

Lasso is preferred when:

* The dataset has many features, and some are expected to be irrelevant.
* Feature selection and model interpretability are important.
* There is a need to reduce dimensionality automatically.

Section 5: Logistic Regression

Q11. Explain the purpose of Logistic Regression.

Answer:

Logistic Regression is used for binary (and multiclass) classification. It models the probability that a given input belongs to a particular class using the sigmoid function to map predictions to a probability between 0 and 1.

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

Answer:

Linear Regression uses Mean Squared Error (MSE) as the cost function.

Logistic Regression uses Log Loss (also called Binary Cross-Entropy) since it is designed for classification tasks and outputs probabilities.

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 that shows the number of **true positives, false positives, true negatives,** and **false negatives.**It helps evaluate classification performance beyond accuracy, allowing calculation of precision, recall, F1-score, etc., which are critical for imbalanced datasets.