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 model learns not only the underlying patterns in the training data but also the noise, resulting in excellent performance on training data but poor generalization to new, unseen data.

•Underfitting happens when a model is too simple to capture the underlying structure of the data, 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 low accuracy on validation/test data.

•Large gap between training and validation error.

•Model performs well on known data but poorly on new, unseen data.

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

•Using regularization techniques (L1, L2).

•Reducing model complexity.

•Collecting more training data.

•Using cross-validation.

•Pruning decision trees or using 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 term to the loss function, discouraging complex models and large coefficients. It helps models generalize better to unseen data.

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

•L1 Regularization (Lasso): Adds the absolute value of coefficients as a penalty term to the loss function. Can shrink some coefficients to exactly zero, thus performing feature selection.

•L2 Regularization (Ridge): Adds the squared value of coefficients as a penalty term. Shrinks coefficients towards zero but rarely makes them exactly zero.

Q6. How does regularization affect the bias-variance tradeoff? Answer: Regularization increases bias but decreases variance. By penalizing large coefficients, it simplifies the model, reducing the risk of overfitting (high variance) but possibly increasing underfitting (high bias).

Section 3: Ridge Regression

Q7. What is Ridge Regression, and how does it differ from standard linear regression? Answer:

Ridge Regression is a type of linear regression that includes an L2 penalty term (sum of squared coefficients) in the loss function. This penalty discourages large coefficients, helping to prevent overfitting. Standard linear regression does not include this penalty.

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

The regularization parameter (λ) controls the strength of the penalty.

If λ = 0, Ridge Regression becomes standard linear regression.

As λ increases, coefficients are shrunk more towards zero, increasing bias but reducing variance and overfitting.

Section 4: Lasso Regression

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

Answer:

Lasso Regression is a linear regression technique that uses

L1 regularization (sum of absolute values of coefficients) as

a penalty. It can shrink some coefficients to exactly zero,

effectively performing feature selection by removing less

important features.

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

Answer:

Lasso Regression is preferred when only a few features are

important and you want automatic feature selection. Ridge

Regression is better when all features are expected to

contribute a little.

Section 5: Logistic Regression

Q11. Explain the purpose of Logistic Regression.

Answer:

Logistic Regression is used for binary or multiclass

classification problems. It models the probability that

a given input belongs to a particular class using the

logistic (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 the

cost function.

•Logistic Regression: Uses Log Loss (Binary Cross-Entropy)

as the cost function, which penalizes incorrect

classifications more heavily and is suitable for probability

outputs.

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

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

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. It shows the counts of true positives, true negatives, false positives, and false negatives.

Significance:

•Helps to calculate metrics like accuracy, precision, recall, and F1-score.

•Provides insight into types of errors the model is making.

•Useful for imbalanced datasets where accuracy alone can be misleading.