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:** When a model learns the training data—including noise and details—too well, causing poor performance on new, unseen data.

**Underfitting:** When a model is too simple to capture the underlying pattern in the data, leading to poor performance on both training and new data.

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

1. High accuracy on training data but low accuracy on test/validation data
2. Large gap between training and test errors.
3. Model performs well on known data but poorly on new/unseen data

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

1-Regularization (L1, L2)

2-Cross-validation

3-Early stopping

4-Simplifying the model (fewer features, smaller architecture)

5-Increasing training data

Section 2: Regularization

Q4. What is regularization, and why is it used in machine learning models? Answer: Regularization is a technique that adds a penalty to a model’s complexity (usually on the size of coefficients) to prevent overfitting.

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

**L1:** Penalizes absolute values; can zero out coefficients (feature selection).

**L2:** Penalizes squares; shrinks coefficients but keeps them nonzero.

Q6. How does regularization affect the bias-variance tradeoff? Answer: Regularization **increases bias** (simplifies the model) but **reduces variance** (less overfitting), helping achieve a better bias-variance balance.

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 adds **L2 regularization** to the cost function.

Q8. What is the effect of the regularization parameter (λ) in Ridge Regression? Answer: The regularization parameter lambda λ in Ridge Regression controls the **strength of the penalty** on the coefficients:

1)- Large 𝜆: More shrinkage of coefficients toward zero - simpler model, higher bias, lower variance.

2)- Small 𝜆: Less shrinkage - model behaves like standard linear regression, lower bias, higher variance.

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**, adding the **absolute values** of the coefficients as a penalty to the cost function.

Lasso can shrink some coefficients **exactly to zero**, effectively **removing irrelevant features** from the model.

his makes Lasso useful for **sparse models** and **automatic feature selection**.

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

Answer:

1-Only a few features are important

2-You need feature selection

3-You want a sparse, interpretable model

4-Data has many features (possibly > samples)

Section 5: Logistic Regression

Q11. Explain the purpose of Logistic Regression.

Answer: **Logistic Regression** is used to **classify data into two or more categories**, typically for **binary classification**

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

Answer: The **cost function in Logistic Regression** is different because it’s designed for **classification**, not regression.

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.

Ans-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.

Ans - B

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

Answer:

**confusion matrix** is a table used to **evaluate the performance** of classification models like **Logistic Regression** by comparing predicted and actual class labels.