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 both patterns and the noise in the training data resulting in poor generalization to unseen data**

**Underfitting : When a model is too simple to capture the underlying patterns in 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 test data**

* **Large gap between training and validation performance**

Q3. What are some common methods to prevent overfitting? Answer: **- Use cross validation**

* **Use regularization**
* **Increase training data**
* **Use drop out**

Section 2: Regularization

Q4. What is regularization, and why is it used in machine learning models? Answer: **It adds a penalty to the loss function to discourse complex models with large coefficient. It helps model to reduce overfitting by simplifying the model**

Q5. Explain the difference between L1 and L2 regularization. Answer**: L1: Adds the sum of the absolute value of the coefficient of the loss function. It shrinks some coefficients to zero**

**L2 : Adds sum of the squares of the coefficients of the loss function .It shrinks coefficients but doenot set them to zero.**

Q6. How does regularization affect the bias-variance tradeoff? Answer**: It increases bias slightly but significantly reduces variance , which helps achieve better generalization on unseen data .**

Section 3: Ridge Regression

Q7. What is Ridge Regression, and how does it differ from standard linear regression? **Answer:** **It is a type of linear regression that includes L2 regularization. It differ from standard linear regression by penalizing large coefficients to reduce overfitting.**

Q8. What is the effect of the regularization parameter (λ) in Ridge Regression? Answer: **The regularization parameter λ controls the strength of the penalty. A larger λ leads to smaller coefficient values, where a smaller λ allows the model to fit more closely to the training data.**

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 exactly to zero, effectively removing those features from the model and performing feature selection.**

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

Answer**: Lasso is performed wen we expect that only few features are relevant. It is useful for sparse models where features selection is important.**

Section 5: Logistic Regression

Q11. Explain the purpose of Logistic Regression.

Answer: **It is used for binary classification problems . It estimates the probability that a given input belongs to a particular class using the logistics function**

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

Answer: **The cost function is the binary class entropy .It is specifically designed to measure the difference between predicted probabilities and actual class labels.**

Section 6: Multiple Choice Questions

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

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.

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

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.

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 a classification model’s performance beyond accuracy , including precision , recall and F1 score .**