ASSIGNMENT : MACHINE LEARNING

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

A) Least Square Error B) Maximum Likelihood C) Logarithmic Loss D) Both A and B

ANS - D) Both A and B

2. Which of the following statement is true about outliers in linear regression?

A) Linear regression is sensitive to outliers
B) linear regression is not sensitive to outliers
C) Can't say D) none of these

ANS - A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?

A) Positive B) Negative C) Zero D) Undefined

ANS - B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?

A) Regression B) Correlation C) Both of them D) None of these

ANS - B) Correlation

5. Which of the following is the reason for over fitting condition?

A) High bias and high variance B) Low bias and low variance
C) Low bias and high variance D) none of these

ANS - C) Low bias and high variance

6. If output involves label then that model is called as:

A) Descriptive model B) Predictive modal C) Reinforcement learning D) All of the above

ANS - B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

A) Cross validation B) Removing outliers C) SMOTE D) Regularization

ANS - D) Regularization

8. To overcome with imbalance dataset which technique can be used?

A) Cross validation B) Regularization C) Kernel D) SMOTE

ANS - D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

A) TPR and FPR B) Sensitivity and precision

C) Sensitivity and Specificity D) Recall and precision

ANS - A) TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

A) True B) False

ANS – B) False

11. Pick the feature extraction from below:

A) Construction bag of words from a email B) Apply PCA to project high dimensional data

C) Removing stop words D) Forward Selection

ANS - A) Construction bag of words from a email

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

A) We don't have to choose the learning rate.

B) It becomes slow when number of features is very large.

C) We need to iterate.

D) It does not make use of dependent variable.

ANS – option A) We don't have to choose the learning rate & option D) It does not make use of dependent variable ,both options are correct.

13. Explain the term regularization ?

ANS - The term regularization refers to a technique in machine learning model, which is used to reduce errors by fitting the function on the given dataset and preventing overfitting and it also enhances the generalization of a model . Regularization adds a penalty term to the model's objective function ,which discourages overly complex models .

Overfitting : This phenomenon occurs when a model learns a data very well and capturing noise but when it comes to apply it perform poorly on new data or unseen data .

Types of regularization techniques :

There are three types of regularization techniques used in machine learning –

1. L1 Regularization – Lasso Regularization
2. L2 Regularization – Ridge Regularization
3. Elastic Net Regularization – L1 and L2 Regularization

1) **L1 Regularization :**

- It is also known as "**Lasso Regularization**" . Full form of **LASSO** is **Least Absolute Shrinkage and Selection Operator .**
- **It adds the absolute values of the coefficient to the objective function .**

2) **L2 Regularization :**

- It is also known as "**Ridge Regularization**" .
- **It adds the "squares value" of the coefficient to the objectives function .**

3) L1 and L2 Regularization :

- It is also known as "**Elastic Net Regularization**" .
- This is the combination of both L1 and L2 Regularization techniques in the objective function .

Q-14) Which particular algorithms are used for regularization ?
ANS – Regularization techniques is applied to different kind of machine learning models to avoiding overfitting and improve the generalization of models.

The common algorithms that can benefit from regularization include :

1. **Linear Regression:**
    - **Algorithms:** Lasso Regression, Ridge Regression, Elastic Net
    - **Explanation:** Regularization helps control the coefficients , preventing overfitting and enhances the generalization of model in linear regression model .
2. **Logistic Regression:**
    - **Algorithms:** Lasso Regression, Ridge Regression, Elastic Net
    - **Explanation:** Regularization helps control the coefficients , preventing overfitting and enhances the generalization of model in logistic regression models.
3. **Neural Networks:**
    - **Algorithm:** Regularization techniques like Dropout, L1/L2 regularization
    - **Explanation:** Regularization can be applied to the weights in neural networks using techniques like Dropout, L1, or L2 regularization.
4. **Decision Trees:**
    - **Algorithm:** Pruning
    - **Explanation:** Regularization in decision trees is often achieved through pruning, removing branches that do not contribute significantly to predictive accuracy.
5. **Random Forest:**
    - **Algorithm:** Pruning, L1/L2 regularization on individual trees
    - **Explanation:** Random forests can use both pruning and regularization on individual trees to control their complexity.

Q-15 Explain the term error present in linear regression equation ?

ANS – The term “error” in linear regression equation specify to difference between the predicted values by models and the actual values in dataset.
> This difference is also known as “residual” .

The representation for linear regression equation is :

$$Y = \beta 0 + \beta 1 X + \varepsilon$$

- Y is the predicted value of the dependent variable(y) for any given value of independent variable(x)
- $\beta 0$ is the intercept , the predicted value of y , when the x = 0
- $\beta 1$ is the regression coefficient , how much we except y to changes as x increases
- X is the independent variable
- ε is the error