MACHINE LEARNING assignment

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- A) Least Square Error

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 model

7. Lasso and Ridge regression techniques belong to _____?

A) Cross validation B) Removing outliers

C) SMOTE D) Regularization

Ans- D) Regularization (L1 and L2 regularization)

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

A) Cross validation B) Regularization

C) Kernel D) SMOTE

Ans- D) SMOTE (Synthetic Minority Over-sampling Technique)

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- B) Apply PCA to project high dimensional data

Description- Principal Component Analysis (PCA) is a feature extraction technique used to reduce the dimensionality of data while retaining as much variance as possible.

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-

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

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

13. Explain the term regularization?

Ans- Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it. Regularization is a technique used to prevent overfitting by adding a penalty term to the model's objective function during training. The objective is to discourage the model from fitting the training data too closely and promote simpler models that generalize better to unseen data. Regularization methods control the complexity of models by penalizing large coefficients or by selecting a subset of features, thus helping to strike the right balance between bias and variance.

Roles of regularization-

- Complexity Control: Regularization helps control model complexity by preventing overfitting to training data, resulting in better generalization to new data.
- Preventing Overfitting: One way to prevent overfitting is to use regularization, which penalizes large coefficients and constrains their magnitudes, thereby preventing a model from becoming overly complex and memorizing the training data instead of learning its underlying patterns.
- Balancing Bias and Variance: Regularization can help balance the trade-off between model bias (underfitting) and model variance (overfitting) in machine learning, which leads to improved performance.
- Feature Selection: Some regularization methods, such as L1 regularization (Lasso), promote sparse solutions that drive some feature coefficients to zero. This automatically selects important features while excluding less important ones.
- Handling Multicollinearity: When features are highly correlated (multicollinearity), regularization can stabilize the model by reducing coefficient sensitivity to small data changes.
- Generalization: Regularized models learn underlying patterns of data for better generalization to new data, instead of memorizing specific examples.

In the field of machine learning, overfitting and underfitting are two critical concepts that directly impact the performance and reliability of models. Overfitting occurs when a model captures noise and patterns specific to the training data, leading to poor generalization on unseen data. On the other hand, underfitting arises when a model is too simple to capture the underlying patterns in the data, resulting in poor performance on both the training and testing datasets.

Regularization plays a pivotal role in enhancing the generalization ability of machine learning models. By mitigating overfitting, regularization techniques improve the model's performance on unseen data, leading to more reliable predictions in real-world scenarios. Additionally, regularization facilitates feature selection and helps in building interpretable models by identifying the most relevant features for prediction.

14. Which particular algorithms are used for regularization?

Ans- Types of Regularization:

L1 Regularization (Lasso):

- Lasso regression is another regularization technique to reduce the complexity of the model. It stands for Least Absolute and Selection Operator.
- It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights.
- Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0.
- It is also called as L1 regularization.
- Some of the features in this technique are completely neglected for model evaluation.
- Hence, the Lasso regression can help us to reduce the overfitting in the model as well as the feature selection.

L2 Regression (Ridge)

- Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions.
- Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as L2 regularization.
- In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called Ridge Regression penalty. We can calculate it by multiplying with the lambda to the squared weight of each individual feature.
- In the above equation, the penalty term regularizes the coefficients of the model, and hence ridge regression reduces the amplitudes of the coefficients that decreases the complexity of the model.
- As we can see from the above equation, if the values of λ tend to zero, the equation becomes the cost function of the linear regression model. Hence, for the minimum value of λ, the model will resemble the linear regression model.
- A general linear or polynomial regression will fail if there is high collinearity between the independent variables, so to solve such problems, Ridge regression can be used.
- It helps to solve the problems if we have more parameters than samples.

Key Difference between Ridge Regression and Lasso Regression

- Ridge regression is mostly used to reduce the overfitting in the model, and it includes all the features present in the model. It reduces the complexity of the model by shrinking the coefficients.
- Lasso regression helps to reduce the overfitting in the model as well as feature selection.

Elastic Net: Combines L1 and L2 regularization, incorporating both the absolute and squared values of the coefficients.

15. Explain the term error present in linear regression equation?

Ans- In a linear regression equation, the term **error** (often referred to as the **residual** or **error term**) represents the difference between the observed value of the dependent variable and the value predicted by the regression model. It accounts for the variation in the dependent variable that cannot be explained by the linear relationship with the independent variables. Here's a detailed explanation:

$Y=\alpha X+\beta\rho+\epsilon$

where:

α,β=Constant parameters

X,ρ=Independent variables

ϵ=Error term

When the actual Y differs from the expected or predicted Y in the model during an empirical test, then the error term does not equal 0, which means there are other factors that influence Y.

Although the error term and residual are often used synonymously, there is an important formal difference. An error term is generally unobservable and a residual is observable and calculable, making it much easier to quantify and visualize. In effect, while an error term represents the way observed data differs from the actual population, a residual represents the way observed data differs from sample population data.