MACHINE LEARNING

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

ANS D) Both A and B

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

ANS A) Linear regression is sensitive to outliers

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

ANS B) Negative

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

ANS C) Both of them

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

ANS C) Low bias and high variance

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

   ANS Predictive modal

7. Lasso and Ridge regression techniques belong to _________?

ANS D) Regularization

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

ANS D) SMOTE

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

ANS A) TPR and FPR

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

ANS B) False

11. Pick the feature extraction from below:

ANS B) Apply PCA to project high dimensional data

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

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

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

13. Explain the term regularization?

ANS Regularization is a technique used in machine learning to prevent overfitting and improve the generalization of a model. Overfitting occurs when a model learns the training data too well, capturing noise and producing a complex model that doesn't generalize well to new, unseen data Regularization introduces a penalty term to the model's objective function, discouraging overly complex models. The two commonly used forms of regularization in linear regression are L1 regularization (Lasso) and L2 regularization (Ridge):

L1 regularization (Lasso): In L1 regularization, the penalty term is the absolute value of the coefficients' sum. It tends to enforce sparsity in the model, encouraging some coefficients to become exactly zero. This can be useful for feature selection.

In L2 regularization, the penalty term is the square of the coefficients' sum. It tends to shrink the coefficients toward zero without necessarily making them exactly zero. Ridge regularization is effective in handling multicollinearity.

14. Which particular algorithms are used for regularization?

1) L1 Regularization (Lasso): In Lasso regression, the regularization term is the absolute value of the coefficients multiplied by a regularization parameter (alpha). It helps in producing sparse models by encouraging some of the coefficients to become exactly zero.
2) L2 Regularization Ridge Regression: Ridge regression adds the squared magnitude of the coefficients to the cost function, multiplied by a regularization parameter (alpha). It penalizes large coefficients but does not typically lead to sparsity in the coefficients.
3) Elastic Net: Elastic Net combines both L1 and L2 regularization terms. It introduces two parameters, alpha and l1_ratio, allowing for a mix of L1 and L2 regularization. This method is useful when there are correlated features.
4) Dropout: Dropout is a regularization technique commonly used in neural networks. During training, random nodes (neurons) are dropped out of the network, which helps prevent co-adaptation of feature detectors and reduces overfitting.
5) Early Stopping: While not a traditional regularization method, early stopping is a technique used to prevent overfitting by monitoring the model's performance on a validation set and stopping the training process when the performance starts to degrade.
6) Weight Decay: Weight decay, also known as L2 regularization in the context of neural networks, adds a penalty term to the loss function based on the squared magnitude of the weights. It discourages the weights from becoming too large.
7) Batch Normalization: a technique that normalizes the inputs of each layer in a neural network, which can act as a form of regularization by reducing internal covariate shift. It helps stabilize and speed up the training process.
8) Data Augmentation: While not a regularization algorithm in the traditional sense, data augmentation involves applying random transformations to the training data (e.g., rotation, scaling, flipping) to increase the diversity of the dataset and improve the model's generalization.

These regularization techniques can be applied in various combinations depending on the specific characteristics of the data and the model being used. The choice of regularization method often depends on the problem at hand and the nature of the data.

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

ANS: In the context of linear regression, the term "error" refers to the difference between the predicted values of the dependent variable (output) and the actual observed values. These errors, also known as residuals, are essentially the discrepancies or mistakes made by the model when trying to estimate the relationship between the independent variables (features) and the dependent variable.

The linear regression equation is typically expressed as:

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n+\varepsilon$$

Y is the dependent variable (the variable we are trying to predict).

$\beta_0$is the y-intercept (the value of Y when all X values are 0).

$\beta_1X_1+\beta_2X_2+\ldots+\beta_n$ are the coefficients associated with the independent variable.

X_1, X_2... X_n representing the strength and direction of the relationship between each independent variable and the dependent variable

ε is the error term, representing the unobserved factors that affect the dependent variable but are not explicitly accounted for by the model. These factors can include measurement errors, omitted variables, or other sources of variability

The linear regression model aims to minimize the sum of squared errors (SSE) or residuals, which is the sum of the squared differences between the predicted and observed values. Mathematically, the sum of squared errors is given by:

$$SSE=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$$

Where Y_i is the observed value of the dependent variable for the i -th data point.

$\hat{Y}_i$ is the predicted value of the dependent variable for the i-th data point based on the regression equation