MACHINE LEARNING In Q1 to Q11, only one option is correct, choose the correct option:

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

A) Least Square Error

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

A) Linear regression is sensitive to outliers

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

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

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

C) Low bias and high variance

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

B) Predictive model

7. Lasso and Ridge regression techniques belong to _____?

D) Regularization

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

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

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

B) False

11. Pick the feature extraction from below:

B) Apply PCA to project high dimensional data

Q12, more than one options are correct, choose all the correct options:

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.

ASSIGNMENT – 39 MACHINE LEARNING Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Regularization is a technique used in machine learning to prevent overfitting and improve the generalization of models. In essence, it involves adding a penalty term to the loss function that a model is trying to minimize. The penalty term is designed to discourage the model from becoming too complex or fitting the training data too closely, which can cause it to perform poorly on new, unseen data.

There are several types of regularization techniques, including:

1. L1 Regularization (Lasso Regression): This technique adds a penalty term proportional to the absolute value of the weights in the model. It tends to produce sparse models, where many of the weights are set to zero.
2. L2 Regularization (Ridge Regression): This technique adds a penalty term proportional to the square of the weights in the model. It tends to produce models with small weights.
3. Dropout Regularization: This technique randomly drops out (i.e., sets to zero) some of the neurons in a neural network during training. This helps prevent the network from becoming too reliant on any one set of features.
4. Early Stopping: This technique stops the training process before the model has fully converged to the training data. This can help prevent overfitting by finding the optimal point where the model has learned the general patterns in the data, but has not yet started to fit the noise.

Overall, regularization is an important technique for improving the performance of machine learning models, especially in cases where the data is limited or noisy.

14. Which particular algorithms are used for regularization?

Regularization can be applied to a wide range of machine learning algorithms, including linear regression, logistic regression, support vector machines, neural networks, and decision trees, among others. Some of the most popular regularization algorithms include:

1. L1 and L2 Regularization: These techniques are commonly used with linear regression and logistic regression.
2. Ridge Regression: This is a type of L2 regularization that is often used with linear regression.
3. Lasso Regression: This is a type of L1 regularization that is often used with linear regression.
4. Elastic Net: This is a combination of L1 and L2 regularization that is often used with linear regression.
5. Dropout: This is a regularization technique that is often used with neural networks.
6. Early Stopping: This is a regularization technique that is often used with neural networks.
7. Tree Pruning: This is a regularization technique that is often used with decision trees.

The choice of which regularization algorithm to use depends on the specific problem at hand and the characteristics of the data being used. In general, L1 and L2 regularization are widely

applicable and effective, while other techniques like dropout and tree pruning are more specialized to certain types of models.

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

In linear regression, the error term refers to the difference between the predicted values of the dependent variable (y) and the actual values of the dependent variable. The goal of linear regression is to minimize the sum of the squared errors (SSE) between the predicted values and the actual values.

The linear regression equation can be written as:

$y = b0 + b1*x + e$

where y is the dependent variable, x is the independent variable, b0 and b1 are the regression coefficients, and e is the error term.

The error term represents the variability in y that is not explained by the independent variable x. It includes all of the factors that affect y but are not included in the model. These could be random factors or unobserved variables that are not accounted for in the model.

By minimizing the sum of the squared errors, linear regression tries to find the line of best fit that represents the relationship between x and y. The line of best fit is the one that minimizes the SSE, which means that it is the line that comes closest to predicting the actual values of y.

In summary, the error term in linear regression represents the part of the dependent variable that is not explained by the independent variable, and minimizing the sum of the squared errors is the goal of linear regression.