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
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:- A) Positive

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:- A) Cross validation

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:- C) Sensitivity and Specificity

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:- C) Removing stop words

In 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.
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 number of features is very large.

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Ans :-
Regularization :-

Sometimes the machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data. This happens because the model is trying too hard to capture the noise present in the training dataset. Noise means the data points that don't really represent the true properties of the data, but random chance. Learning such data points, makes model more flexible, at the risk of overfitting and is known as over-fitted.

Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid over fitting. In other words, this technique converts a complex model into a simpler one, so as to avoid the risk of overfitting and shrinks the coefficients, for lesser computational cost.

14. Which particular algorithms are used for regularization?
Ans:-

The Algorithms used for Regularization are as follows.
1. Ridge Regression (L2 regularization or L-2 norm)
2. Lasso Regression (L1 regularization or L-1 norm)
3. Elastic-Net Regression

Ridge Regression:-

Ridge regression is one of the types of linear regression in which we introduce a small amount of bias, known as Ridge regression penalty so that we can get better long-term predictions.

The penalty term regularizes the coefficients of the model, and hence ridge regression reduces the magnitudes of the coefficients that help to decrease the complexity of the model.

Usage of Ridge Regression:

- When we have the independent variables which are having high collinearity (problem of multicollinearity) between them, at that time general linear or polynomial regression will fail so to solve such problems, Ridge regression can be used.
- If we have more parameters than the samples, then Ridge regression helps to solve the problems.

Limitation of Ridge Regression:

- Not helps in Feature Selection: It decreases the complexity of a model but does not reduce the number of independent variables since it never leads to a coefficient being zero rather only minimizes it. Hence, this technique is not good for feature selection.
- Model Interpretability: Its disadvantage is model interpretability since it will shrink the coefficients for least important predictors, very close to zero but it will never make them exactly zero. In other words, the final model will include all the independent variables, also known as predictors.

Lasso Regression

Lasso regression is another variant of the regularization technique used to reduce the complexity of the model. It stands for Least Absolute and Selection Operator.

Lasso regression is another variant of the regularization technique used 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 includes the absolute weights instead of a square of weights.

The L1 penalty has the effect of forcing some of the coefficient estimates to be exactly equal to zero which means there is a complete removal of some of the features for model evaluation when the tuning parameter λ is sufficiently large. Therefore, the lasso method also performs Feature selection and is said to yield sparse models.

Limitation of Lasso Regression:

- Problems with some types of Dataset: If the number of predictors is greater than the number of data points, Lasso will pick at most n predictors as non-zero, even if all predictors are relevant.
- Multi-collinearity Problem: If there are two or more highly collinear variables, the LASSO regression selects one of them randomly which is not good for the interpretation of the model.

Elastic-Net Regression

Elastic-Net is a regularized regression method that linearly combines the L1 and L2 penalties of the LASSO and Ridge methods respectively.

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

Ans:-
Term error present in linear regression equation

The normal linear regression model involves finding the best fitting linear model for observed data that shows the relationship between two variables.

When we plot a data on a graph it is plotted as a scatter plot. That means not all the data points lie on the line represented by our linear regression equation. But we want our linear regression equation like $y = \beta_0 + \beta_1 x$ to be the best fit. Best fit means we want the distance from each point to the line to be minimum possible. This distance from each point to our equation line is known as the Error Term. Hence in best fit we want this Error Term to be minimized to the lowest possible value.

So we can write our function as $y = \beta_0 + \beta_1 x + \varepsilon$ where β_0 and β_1 are constants and ε is an error term.