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

   Answer: 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

   Answer: 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

   Answer: 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

   Answer: C) Both of them

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

   Answer: 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

   Answer: B) Predictive model

7. Lasso and Ridge regression techniques belong to _____?
   A) Cross validation B) Removing outliers
   C) SMOTE D) Regularization

   Answer: D) Regularization

8. To overcome with imbalance dataset which technique can be used?
   A) Cross validation B) Regularization
   C) Kernel D) SMOTE

   Answer: D) SMOTE (Systematic Minority Oversampling 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

MACHINE LEARNING

Answer: 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

    Answer: A) True

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

    Answer: B) Apply PCA to project high dimensional data

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.

    Answer: A) We don't have to choose the learning rate
    B) It becomes slow when number of features is very large

MACHINE LEARNING

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

13. Explain the term regularization?
14. Which particular algorithms are used for regularization?
15. Explain the term error present in linear regression equation?

13) **Regularization:**

- Regularization technique is used in the machine learning model in order to minimize the adjusted loss function and **prevent Overfitting and Underfitting**.
- It helps to choose a simple model rather than a complex model.
- Regularization works by shrinking the beta coefficients of a linear regression model
- For linear regression, the regularization has two terms in the loss function:

  1. The Ordinary Least Squares (OLS) function, and
  2. The penalty term.

- It helps prevent the problem of overfitting, makes the model more robust, and decreases the complexity of a model.
- The regularization techniques are **Lasso Regression, Ridge Regression, and Elastic Net Regression.**
- It imposes a higher penalty on the variable having higher values, and hence, it controls the strength of the penalty term.
- Both L1 and L2 can add a penalty to the cost depending upon the model complexity, so at the place of computing the cost by using a loss function
- In order to prevent overfitting, regularization is most-approaches mathematical technique, it achieves this by panelizing the complex ML models via adding regularization terms to the loss function/cost function of the model.
- Regularization appends penalties to more complex models and arranges potential models from slightest overfit to greatest.

MACHINE LEARNING

14) Commonly Used Algorithm in Regularization:

1. Lasso (L1) Regularization
2. Ridge (L2) Regularization
3. Elastic Net

Lasso (L1) Regularization:

- It modifies the Over fitted or under fitted model by adding the penalty equivalent to the sum of the absolute values of Coefficient.
- Lasso Regularization perform Coefficient Minimization. Instead of squaring the magnitude of coefficient, it takes the true value of coefficient.
- Reducing the value of coefficient to the lowest.
- Least Absolute Shrinkage and Selection Operator (or LASSO) Regression penalizes the coefficients to the extent that it becomes zero.
- It eliminates the insignificant independent variables.
- It adds an L1 penalty that is equal to the absolute value of the magnitude of coefficient, or simply restricting the size of coefficients.
- L1 regularization is relatively more expensive in computation, it can't be solved in the context of matrix measurement and heavily relies on approximations.
- L1 regularization can be helpful in features selection by eradicating the unimportant features,

Ridge(L2) Regularization:

- In Ridge(L2) Regularization it modifies the Over fitted or under fitted by adding the penalty equivalent to sum of squares of magnitude of coefficient.
- It is used to prevent Multicollinearity.
- The Ridge regression technique is used to analyze the model where the variables may be having multicollinearity. It reduces the insignificant independent variables though it does not remove them completely.
- It uses the L2-norm as the penalty.
- It adds an L2 penalty which is equal to the square of the magnitude of coefficients. For example, Ridge regression and SVM implement this method.

MACHINE LEARNING

- It gives more accurate predictions when the output variable is the function of whole input variables.
- Ridge Regression also works when we have Discrete variables like high fat, low fat, etc.
- Ridge Regression helps reduce variance by shrinking the parameters and making our predictions less sensitive to them. The penalty term contains all the parameters except for the "y prediction".

Elastic Net:

- When L1 and L2 regularization combine together, it becomes the elastic net method, it adds a hyperparameter.
- The model using elastic net regression allows the learning of the sparse model where some of the points are zero, similar to Lasso regularization, and yet maintains the Ridge regression properties.

L1: $$R(\theta)=\|\theta\|_1=\sum_{i=1}^{n}|\theta_i|$$

L2: $$R(\theta)=\|\theta\|_2^2=\sum_{i=1}^{n}\theta_i^2$$

MACHINE LEARNING

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

- An error term is a **residual variable** produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables.
- An error term appears in a statistical model, like a regression model, to indicate the uncertainty in the model. The error term is a residual variable that accounts for a lack of perfect goodness of fit.
- The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters *e*, ε, or u.
- Linear regression is a form of analysis that relates to current trends experienced by a particular security or index by providing a relationship between a dependent and independent variable.
- The error term in a regression equation represents the effect of the variables that were omitted from the equation.

(i) the combined effect of the omitted variables is independent of each variable included in the equation.

(ii) the combined effect of the omitted variables is independent across subjects.

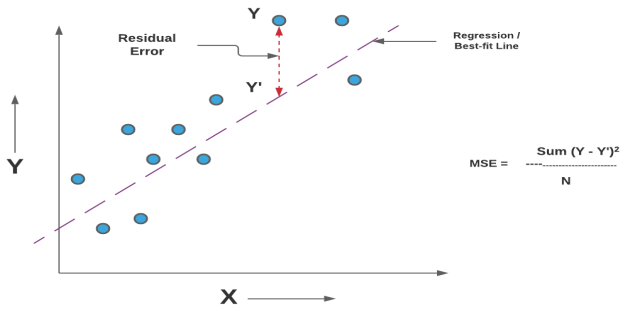
(iii) the combined effect of the omitted variables has expectation 0.

Linear Regression Equation is

- **Y=a+bX**

- Model is linear in terms of coefficient and error term.
- Independent variables are uncorrelated with Residual term.
  - **r=y-(mx+b)**
- Predicting a quantitative response using single feature (“Input variable”)
- If the plot shows a straight line without any deviation, which means the error is normally distributed. The linear regression model assumes no autocorrelation in error terms. If there will be any correlation in the error term, then it will drastically reduce the accuracy of the model.
- The error term (μ i) is a random real number, μ may assume any positive, negative or zero value upon chance

MACHINE LEARNING



- The dependent and independent variables show a linear relationship between the slope and the intercept.
- The value of the residual (error) is not correlated across all observations.
- The **residual standard deviation (or residual standard error) is a measure used to assess how well a linear regression model fits the data**. (The other measure to assess this goodness of fit is R2).
- Linear Regression: Residual Standard Error *is used to evaluate linear regression* goodness of fit *by estimating its residual standard deviation.*
- The error is a random variable with a mean of zero conditional on the explanatory variables.