Machine Learning

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)High Variance and Low Bias

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

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

Q13 and Q15 are subjective answer type questions, Answer them briefly.
13. Explain the term regularization?

Regularization is tuning or selecting the preferred level of model complexity so that your models are better at predicting . It refers to the techniques that are used to calibrate the machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

14. Which particular algorithms are used for regularization?

LASSO(L1), Ridge(L2) and Drop out are the algorithms used for regularization.

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

In a regression equation, the error term (also known as the residual point) represents the variability in the dependent variable that is not explained by the independent variables included in the model. It captures all other factors that influence the dependent variable but are not explicitly accounted for in the regression model.