MachineLearning_worksheet

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
Answer1-: 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
Answer2-: 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
Answer3-: C, Zero

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
Answer4-: 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
Anwer5-: C, Low bias and high variance

6. If output involves label then that model is called as:
A) Descriptive model B) Predictive model
C) Reinforcement learning D) All of the above
Answer6-: 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

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
Answer-:A, TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
A) True B) False
Answer:-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
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.
Answer12-: A,B,D

13. Explain the term regularization?
Answer13:- Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it. 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 by introducing noise in the output, and hence the model is called overfitted. This problem can be deal with the help of a regularization technique.

This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model. It mainly regularizes or reduces the coefficient of features toward zero. In simple words, "In regularization technique, we reduce the magnitude of the features by keeping the same number of features.

14. Which particular algorithms are used for regularization?
Answer14-: There are three types of algorithms which used for regularization.

1.Ridge Regression

2.LASSO (Least Absolute Shrinkage and Selection Operator) Regression

3.Elastic-Net Regression

15. Explain the term error present in linear regression equation?
Answer15-: Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price at a particular time and the price that was actually observed.