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

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 outlier

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 : 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

ANSWER : 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

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

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) TRP and FRP

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

A) True B) False

ANSWER : 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

ANSWER : B) Apply PCA to Project High Dimenstional 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 : B) it become slow when number of features is very large. **&** C) We need to iterate.

13. Explain the term regularization?

ANSWER : In machine learning model, when training data is good but fails to work on testing data and not able to predict the data means there is a problem of overfitting of data there.

So this problem can be solve by using regularization. So regularization is a process to avoid overfitting of the data in training stage so that the machine will predict the data properly.

14. Which particular algorithms are used for regularization?

ANSWER : 1. Ridge Regression and 2. LASSO Regression

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

ANSWER : It is called the Mean Absolute Error. Means How much is difference coming between the actual data and the predicted data.

Its formula is summation of (machine answer – original answer) divide by total number of answer.

If error is less then your machine is giving write answer.