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

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

3. A line falls from left to right if a slope is _____?

A) Positive B) Negative

C) Zero D) Undefined ANS-A

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-C

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

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

7. Lasso and Ridge regression techniques belong to _____?

A) Cross validation B) Removing outliers

C) SMOTE D) Regularization ANS-D

8. To overcome with imbalance dataset which technique can be used?

A) Cross validation B) Regularization

C) Kernel D) SMOTE ANS-D

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-A

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

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

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,C

.

13. Explain the term regularization?

- ANS- Regularization is one of the most important concepts of machine learning. This technique prevents the model from overfitting by adding extra information to it.

- It is a form of regression that shrinks the coefficient estimates towards zero. In other words, this technique forces us not to learn a more complex or flexible model, to avoid the problem of overfitting

14. Which particular algorithms are used for regularization?

ANS-Algorithms that are used for regularization are Ridge Regression, LASSO, Elastic-Net Regression.The working of all these algorithms is quite similar to that of Linear Regression, it's just the loss function that keeps on changing.

Ridge Regression - Ridge regression is a method for analyzing data that suffer from multi-collinearity.

LASSO Regression- LASSO is a regression analysis method that performs both feature selection and regularization in order to enhance the prediction accuracy 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-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. In instances where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero.