ASSIGNMENT – 39

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: 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: A) High 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: A) Descriptive model

7. Lasso and Ridge regression techniques belong to \_\_\_\_\_\_\_\_\_?

A) Cross validation B) Removing outliers

C) SMOTE D) Regularization

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

A) Cross validation B) Regularization

C) Kernel D) SMOTE

Answer: B) Regularization

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.

1. True B) False

Answer: A) True

11. Pick the feature extraction from below:

A) Construction bag of words from an 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 and D

ASSIGNMENT – 39

MACHINE LEARNING

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

13. Explain the term regularization?

Answer:

It 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.
* Now, let’s understand the “How flexibility of a model is represented?”
* For regression problems, the increase in flexibility of a model is represented by an increase in its coefficients, which are calculated from the regression line.
* In simple words, “In the Regularization technique, we reduce the magnitude of the independent variables by keeping the same number of variables”. It maintains accuracy as well as a generalization of the model.

Techniques of Regularization:

Mainly, there are two types of regularization techniques, which are given below:

* Ridge Regression
* Lasso Regression

14. Which particular algorithms are used for regularization?

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

Answer:

Linear regression most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by:

measuring the distance of the observed y-values from the predicted y-values at each value of x;

squaring each of these distances;

calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE.