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

**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) 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

**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: 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

**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) 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**.

**ASSIGNMENT – 39**

**MACHINE LEARNING**

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

**13. Explain the term regularization?**

**Ans:** Regularization is a technique which is used for reduction of the errors by fitting the function appropriately on the given training set as well as avoiding overfitting. It is a type of regression which shrinks the coefficient that estimates towards zero. Simply put, this technique indicates not to learn a more complex form or flexible model, to avoid the problem of overfitting.

**14. Which particular algorithms are used for regularization?**

**Ans:**

* **Ridge Regression**: The function of Ridge regression is that it shrinks the coefficients by helping it to reduce the model complexity and multi-collinearity. It is also termed as L-2 norm.
* **LASSO (Least Absolute Shrinkage and Selection Operator) Regression:** The function of **LASSO regression is that it administers in conversion of coefficients which has less important features to zero, which also helps in feature selection, as well as it shrinks the coefficients of remaining features to assist in reduction of the model complexity to avoid overfitting. It is also termed as L-1 norm.**
* **Elastic-Net Regression:** It is a regularized regression method which linearly combines both the L1 and L2 penalties of the LASSO and Ridge methods.

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

**Ans**: 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 coefficient that results in the smallest MSE.