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?
Answer:) Least Square Error

2. Which of the following statement is true about outliers in linear regression?

Answer:) Linear regression is sensitive to outliers

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

Answer:) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?

Answer:) Correlation

5. Which of the following is the reason for over fitting condition?

Answer:) Low bias and high variance

6. If output involves label then that model is called as:

Answer:) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

Answer:) Regularization

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

Answer:) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

Answer:) TPR and FPR

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

Answer:) False

11. Pick the feature extraction from below:

Answer:) A) Construction bag of words from a email B) Apply PCA to project high dimensional data C) Removing stop words

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?

Answer:) .

B) It becomes slow when number of features is very large.

D) It does not make use of dependent variable.

ASSIGNMENT – 39 MACHINE LEARNING

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

13. Explain the term regularization?

Answer:

When we use regression models to train some data, there is a chance that the model will overfit the given training data set. Regularization helps sort overfitting problem by restricting the degree of freedom of a given equation. i.e. simply reducing the number of degrees of polynomial function by reducing their corresponding weights. In a linear equation we do not want huge coefficients as a small change in weight can make a large difference for the dependant variable (Y). So, the regularization constraints the weight of such features to avoid overfitting.

14. Which particular algorithms are used for regularization?

Answer

: Ridge Regression, Lasso (Least Absolute Shrinkage and Selection Operator) and Elastic-Net Regression.

15. Explain the term error present in linear regression equation

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