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

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

3. A line falls from left to right if a slope is_____?
   A) Positive B) ***Negative***
   C) Zero D) Undefined

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

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

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

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***

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***

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

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.

MACHINE LEARNING

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

13. Explain the term regularization?

Ans) Regularization is a technique used for tuning the function by adding an additional penalty term in the error function.

The additional term controls the excessively fluctuating function such that the coefficients don't take extreme values.

14. Which particular algorithms are used for regularization?

Ans) There are mainly three main regularization techniques, which are

- Ridge Regression. (L2 Norm)
- Lasso Regression (L1 Norm)
- Dropout

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

Ans) An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results.

The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.

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.

This is how the error could be calculated

Answers in red colour