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

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

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

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

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

C) Low bias and high variance

6. If output involves label then that model is called as:
A) Descriptive model B) Predictive model C) Reinforcement learning D) All of the above

B) Predictive model

7. Lasso and Ridge regression techniques belong to _____? A) Cross validation B) Removing outliers C) SMOTE D) Regularization

D) Regularization

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

A) Cross validation B) Regularization C) Kernel D) SMOTE

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

C) Sensitivity and Specificity

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

A) True B) False

B) False

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

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.

A) We don't have to choose the learning rate. B) It becomes slow when number of features is very large

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

13. Explain the term regularization?

Ans) The term 'regularization' refers to a set of techniques that regularizes learning from particular features for traditional algorithms or neurons in the case of neural network algorithms. It normalizes and moderates weights attached to a feature or a neuron so that algorithms do not rely on just a few features or neurons to predict the result. This technique helps to avoid the problem of overfitting. To understand regularization, let's consider a simple case of linear regression. Mathematically, linear regression is stated as below:

$y = w_0 + w_1x_1 + w_2x_2 + ….. + w_nx_n$

where y is the value to be predicted;

$x_1, x_2, …., x_n$ are features that decides the value of y;

w_0 is the bias;

$w_1, w_2, ….., w_n$ are the weights attached to $x_1, x_2, …., x_n$ relatively.

Now to build a model that accurately predicts the *y* value, we need to optimize above mentioned bias and weights. To do so, we need to use a loss function and find optimized parameters using gradient descent algorithms and its variants.

14. Which particular algorithms are used for regularization?

Ans) There are three main regularization techniques, namely:

1. Ridge Regression (L2 Norm)
2. Lasso (L1 Norm)
3. Dropout

15. Explain the term error present in linear regression equation

Ans) 1.An error term appears in a statistical model, like a regression model, to indicate the uncertainty in the model.

2.The error term is a residual variable that accounts for a lack of perfect goodness of fit.

3. Heteroskedastic refers to a condition in which the variance of the residual term, or error term, in a regression model varies widely.