MACHINE LEARNING

Q1 to Q11 have only one correct answer

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 :- A) Regression

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 :- A) Descriptive 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 :- 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

Ans :- A) True

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

Q12, more than one options are correct

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.

Q13 and Q15 are subjective answer type questions.

13) Explain the term regularization?

Ans :- Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it. When we use regression models to train some data, there is a good chance that the model will overfit the given training data set. Regularization helps sort this overfitting problem by restricting the degrees of freedom of a given equation i.e. simply reducing the number of degrees of a polynomial function by reducing their corresponding weights.
In a linear equation, we do not want huge weights/coefficients as a small change in weight can make a large difference for the dependent variable (Y). So, regularization constraints the weights of such features to avoid overfitting.
The regularization techniques are Lasso Regression, Ridge Regression, and Elastic Net Regression. Regularization can be used for feature reduction.

14) Which particular algorithms are used for regularization?

Ans :- There are three main algorithms used for regularization.

1) LAASO
2) RIDGE
3) ELASTIC NET (Less popular)

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

Ans :-