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
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
3. A line falls from left to right if a slope is_____?
   A) Positive B) Negative
   C) Zero D) Undefined Ans:B
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:C
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:A
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
7. Lasso and Ridge regression techniques belong to_____?
   A) Cross validation B) Removing outliers
   C) SMOTE D) Regularization Ans:D
8. To overcome with imbalance dataset which technique can be used?
   A) Cross validation B) Regularization
   C) Kernel D) SMOTE Ans:D
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
10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
    A) True B) False Ans:B
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

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,B,C

MACHINE LEARNING

MACHINE LEARNING

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

13. Explain the term regularization?

14. Which particular algorithms are used for regularization?

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

13Ans)Regularization:- This is a technique used to reduce complexity of model and it is uded to prevent the model from over fitting or noise.

It will allow to maintain variables in the model by reducing magnitude of variables. So it maintains accuracy. Regularization works by adding a penalty or complexity term or shrinkage term with Residual Sum of Squares (RSS) to the complex model.

The concept of balancing bias and variance, is helpful in understanding the phenomenon of overfitting.One of the ways of avoiding is using cross validation,that helps in estimating the error over test set, and in deciding what parameters work best for your model.Regularization , significantly reduces variance of model, without substantial increase in bias.

14Ans)Algorithms used in Regularization are:

1.Lasso Regression

2.Ridge Regression

3.Elastic Net Regression

15ans)Linear Regression uses one independent variable to explain or predict the outcome of dependent variable Y

Linear Regression:Y=a+bx+e

Y is dependant variable

X is independent variable

E is Residual error