MACHINE LEARNING

Q1: 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

Answer: A) Least Square Error

Q2: 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

Answer: A) Linear regression is sensitive to outliers

Q3: A line falls from left to right if a slope is _____?

A) Positive B) Negative C) Zero D) Undefined

Answer: B) Negative

Q4: 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

Answer: B) Correlation

Q5: 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

Answer: C) Low bias and high variance

Q6: If output involves label then that model is called as: A) Descriptive model B) Predictive modal C) Reinforcement learning D) All of the above

Answer: B) Predictive model

Q7: Lasso and Ridge regression techniques belong to _____? A) Cross validation B) Removing outliers C) SMOTE D) Regularization

Answer: D) Regularization

Q8: To overcome with imbalance dataset which technique can be used? A) Cross validation B) Regularization C) Kernel D) SMOTE

Answer: D) SMOTE

Q9: 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

Answer: A) TPR and FPR

Q10: In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less. A) True B) False

Answer: B) False

Q11: 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

Answer: 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.

Answer:B&D

Q13: Explain the term regularization?

Answer: Regularization is a set of methods for reducing overfitting in machine learning models. Typically, regularization trades a marginal decrease in training accuracy for an increase in generalizability. Regularization provides a range of techniques to correct the overfitting issues by restricting the degree freedom in a given equation in machine learning models.

Q14: Which particular algorithms are used for regularization?

Answer: The linear regression models used regularization techniques to come out of overfitting issues. The two important regularization techniques used in linear regression are LASSO and RIDGE.

Q15: Explain the term error present in linear regression equation?

Answer: A term error is a marginal error within the linear regression model .It refers to the sum of deviation within the regression line which gives an explanation for the difference between predicted value and actual value.