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

   Answer: (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

   Answer: (A)

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

   Answer : (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

   Answer: (B)

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

   Answer : (C)

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

   Answer: (B)

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

   Answer : (D)

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

   Answer : (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

MACHINE LEARNING

Answer : (C)

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

Answer: True

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

Answer : (A) , (B) , (C)

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 : (A) , (B)

MACHINE LEARNING

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

13. Explain the term regularization?

Answer: This is a form of regression that constrains / regularizes or shrinks the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of over fitting.

A simple relation for linear regression looks like this

Y ≈ β0 + β1X1 + β2X2 + …+ βpXp.

14. Which particular algorithms are used for regularization?

Answer: There are three main regularization techniques, namely:

1. Ridge Regression (L2 Norm) –

   Ridge regression is also called L2 norm or regularization. When using this technique, we add the sum of weight's square to a loss function and thus create a new loss function which is denoted thus:

$$\text{Loss} = \sum_{j=1}^{m}\left(Yi - Wo - \sum_{i=1}^{n} Wi\,Xji\right)^2 + \lambda\sum_{i=1}^{n} Wi^2$$

2. Lasso (L1 Norm)

This technique is different from ridge regression as it uses absolute weight values for normalization. "**Λ"** is again a tuning parameter and behaves in the same as it does when using ridge regression. As loss function only considers absolute weights, optimization algorithms penalize higher weight values. In ridge regression, loss function along with the optimization algorithm brings parameters near to zero but not actually zero, while lasso eliminates less important features and sets respective weight values to zero. Thus, lasso also performs feature selection along with regularization.

$$\text{Loss} = \sum_{j=1}^{m}\left(Yi - Wo - \sum_{i=1}^{n} Wi\,Xji\right)^2 + \lambda\sum_{i=1}^{n} |Wi|$$

3. Dropout

Dropout is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons. In neural nets, fully connected layers are more prone to over fit on training data. Using dropout, you can drop connections with 1-p probability for each of the specified layers. Where p is called keep probability parameter and which needs to be tuned.

Dropout decreases over fitting by avoiding training all the neurons on the complete training data in one go

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

Answer: The standard error of the regression (S), also known as the standard error of the estimate, represents the average distance that the observed values fall from the regression line. Conveniently, it tells us how wrong the regression model is on average using the units of the response variable. Smaller values are better because it indicates that the observations are closer to the fitted line. The error term includes everything that separates our model from actual reality. This means that it will reflect nonlinearities, unpredictable effects, measurement errors, and omitted variables.