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: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
ANS: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
ANS:C

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:C

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

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

ASSIGNMENT – 39
MACHINE LEARNING
Q13 and Q15 are subjective answer type questions, Answer them briefly.
13. Explain the term regularization?
ANS) The word regularize means to make things regular or acceptable. This is exactly why we use it for. Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting.
14. Which particular algorithms are used for regularization?

There are three main regularization techniques, namely:

- Ridge Regression (L2 Norm)
- Lasso (L1 Norm)
- Dropout

    Ridge regression: It 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:

    The original loss function is modified by adding normalized weights. Here normalized weights are in the form of squares.

    λ is the parameter that needs to be tuned using a cross-validation dataset. When we use $\lambda=0$, it returns the residual sum of square as loss function which you chose initially. For a very high value of λ, loss will ignore core loss function and minimize weight's square and will end up taking the parameters' value as zero.Now the parameters are learned using a modified loss function. To minimize the above function, parameters need to be as small as possible. Thus, L2 norm prevents weights from rising too high.

Lasso Regression (L1 Regularization)

Also called L1 regularization and denoted as below:

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.

Dropout

1. Dropout is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons.
2. In neural nets, fully connected layers are more prone to overfit 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**.

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

An **error term** in statistics is a value which represents how observed data differs from actual [population data](). It can also be a variable which represents how a given statistical model differs from reality. The error term is often written ε.