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?
   Ans: A) Least Square Error
2. Which of the following statement is true about outliers in linear regression?
   Ans: A) Linear regression is sensitive to outliers
3. A line falls from left to right if a slope is _____?
   Ans: A) Positive
4. Which of the following will have symmetric relation between dependent variable and independent variable?
   Ans: C) Both of them (Regression and Correlation)
5. Which of the following is the reason for over fitting condition?
   Ans: A) High bias and high variance
6. If output involves label then that model is called as:
   Ans: B) Predictive modal
7. Lasso and Ridge regression techniques belong to _____?
   Ans: D) Regularization
8. To overcome with imbalance dataset which technique can be used?
   Ans: D) SMOTE (Synthetic Minority Oversampling Technique)
9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?
   Ans: C) Sensitivity and Specificity
10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
    Ans: B) False
11. Pick the feature extraction from below:

Ans: A) Construction bag of words from a email
B) Apply PCA to project high dimensional data
C) Removing stops words

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.

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

13. Explain the term regularization?

Ans: Regularizations are techniques used in machine learning to reduce the error by fitting a function appropriately on the given training set and avoid over fitting and make things regular or acceptable. Regularization is the process which regularizes or shrinks the coefficients towards zero. In simple words, regularization discourages learning a more complex or flexible model, to prevent over fitting.

There are two main regularization techniques, namely Ridge Regression and Lasso Regression. They both differ in the way they assign a penalty to the coefficients.

14. Which particular algorithms are used for regularization?

Ans: There are three main regularization algorithms:

1. Ridge Regression (L2 Norm): Ridge regression is one of the types of linear regression in which we introduce a small amount of bias, known as Ridge regression penalty so that we can get better long-term predictions. In Statistics, it is known as the L-2 norm.

2. Lasso (L1 Norm): Lasso regression is another variant of the regularization technique used to reduce the complexity of the model. It stands for Least Absolute and Selection Operator. It is similar to the Ridge Regression except

that the penalty term includes the absolute weights instead of a square of weights.

3. Dropout: Dropout is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons. With dropout, you are left with a reduced network as dropped out neurons are left out during that training iteration. Dropout decreases over fitting by avoiding training all the neurons on the complete training data in one go. It also improves training speed and learns more robust internal functions that generalize better on unseen data.

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

Ans: Linear regression most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by:

1. measuring the distance of the observed y-values from the predicted y-values at each value of x;
2. squaring each of these distances;
3. Calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient results in the smallest MSE.