

# Machine Learning Assignment-4

## Answers:

1. (c) between -1 and 1
2. (c) Recursive feature elimination
3. (a) Linear
4. (a) Logistic regression
5. (b) Same as old coefficient of "X"
6. (b) Increase
7. (c) Easy to interpret
8. (b) Using Unsupervised learning techniques and C) Linear combinations of linear variables.
9. (a), (b) and (d)
10. (a) max\_depth and (d) min\_samples\_leaf.
11. An outlier is an object that deviates significantly from the rest of the objects. They can be caused by measurement or execution error. The analysis of outlier data is referred to as outlier analysis or outlier mining. In IQR method of outlier detection, we take IQR value as difference between Q3 and Q1 and set up limit at  $(1.5 \times \text{IQR} + Q3)$  and  $(Q1 - 1.5 \times \text{IQR})$ .
12. Bagging:
  - The Main Goal of Bagging is to decrease variance, not bias.
  - In Bagging multiple training data-subsets are drawn randomly with replacement from the original dataset.
- Boosting:
  - The Main Goal of Boosting is to decrease bias, not variance.
  - In Boosting new sub-datasets are drawn randomly with replacement from the weighted (updated) dataset.
13. The adjusted R-squared is a modified version of R-squared that adjusts for the number of predictors in a regression model.  
Calculation of  $R^2 = 1 - [(1 - R^2) \times (n - 1) / (n - k - 1)]$
14. Normalization changes the values at a standard scale, whereas Standardization assumes data on a Gaussian distribution and measures variable at a different scale.
15. Cross-validation is evaluating various learning algorithms by dividing the dataset into parts and comparing results.
  - Pro: More efficient use of data as every observation is used for training and testing.
  - Con: This tends to be expensive as the training algorithm has to run 'N' times from scratch and is time taking.

