

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) Increases
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. Outliers are data points that are significantly different from the remaining dataset and lead to skewness in the data. 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. Main difference between Bagging and Boosting algorithm is Bagging decreases variance but not the bias, whereas Boosting decreases bias but not the variance.
13. It is a modified version of R-squared that takes into account the predictors that aren't significant in regression model. Adjusted r square is calculated by dividing the residual mean square error by total mean square error.
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:
 - i. More efficient use of data as every observation is used for training and testing.
 - Con:
 - i. This tends to be expensive as the training algorithm has to run 'N' times from scratch and is time taking.