Machine Learning

Worksheet-1

1. C

2. C

3. C

4. B

5. C

6. B

7. C

8. B&C

9. B&C

10. A,B&D

11. Outliers are extreme values that differ from other data. Outliers are value that is much smaller or larger than most of the other values in a the data

IQR is calculated as the difference between the 75th and 25th percentiles of the data.We can then calculate the cutoff for outliers as 1.5 times the IQR and subtract this cutoff from 25th percentile and add it to the 75th percentile to give the actual limits on data.

12. Bagging solves the overfitting problem, but boosting increases it as bagging decreases variance but does not decrease bias whereas boosting decreases bias but doesn't decrease variance. In bagging, each model receives equal weight, but in boosting, models are weighted according to their performance.

13.Adjusted R squared in logistic regression is the r squared value which has been computed taking into consideration how much a single variable has effect on the output rather than all the variable as in logistic regression the increase in no. of variable tends the model to overfit and that results in failure in real world data.

14. standardization does not have a bounding range. Standardization transforms data to have a mean of 0 and standard-deviation of 1.

Normalization means to scale a variable to have a values between 0 and 1

15. Cross-validation is a technique for evaluating the models by training several models on folds of the available input data and evaluating them on the folds of the data. Each data part is sent as input and output so that the model doesn't overfit and gives out the best result in real world data. One of the advantage of using cross-validation is, it prevents our model from over-fitting. And disadvantage is that, We split the dataset into multiple folds and train the algorithm on different folds, which increases the training time.