WORKSHEET 4 MACHINE LEARNING

1) C

2) C
3) C
4) D
5) C
6) D
7) C
8) B and C
9) A, B and D
10) A,B and D

11) The abnormal data that is responsible for skewness or in other words the any data point that lies beyond three standard deviation of mean within the distribution is called as an outlier. The inter quartile range (IQR) is the difference between Q3 (75^{th} percentile) and Q1 (25^{th} percentile). By multiplying this IQR 1.5 times, we can find a value which can be subtracted from Q1 to find the lower boundary and added to Q3 to find the upper boundary for the outliers. And thus, any data point that lies above upper boundary or below lower boundary are considered as the outliers.

12) In bagging algorithm weak model learns independently in parallel from each other and then their learning is combined to give an average output, whereas in boosting algorithm weak model learns sequentially which means learning of one model is fed to the next model and so on.

13) The adjusted R^2 is used to identify whether adding any additional predictors would improve the model working or not. In order to calculate

adjusted R^2 , residual mead square error is first calculated and then divided by total mean square error. The output obtained is then subtracted from 1.

14) Standardization is rescaling the distribution value between 0 and 1 and is used when the data is normally distributed, whereas normalization is the process of changing the feature values so that numeric column in dataset have the common scale and it is mostly used when the data is not normally distributed.

15) Cross-validation is a technique that is used to test the model efficiency by training it on subsets of the available input data and then testing on the unseen data.

The advantage of this method is that it by getting more metrics it allows us to determine the consistency of the algorithm.

The disadvantage is that the training algorithm has to rerun k number of times from the beginning.