ANS1-c(between –1 and 1)

ANS2-c

ANS3-c(hyperplane)

ANS4-b

ANS5-c(old coefficient of x / 2.205)

ANS6-b(increases)

ANS7-c

ANS8-B&C

ANS9-C&D

ANS10-A,B&D

ANS 11- Outliers are abnormalities which are distinct from other data points as they tend to make data heavily tailed (skewed).Outliers are present in data due to wrong collection of data or due to extraordinary or under performance of some data points in the dataset .

The standard deviation of the Outliers are beyond the range of –3 to 3.Outliers are removed using zscore which determines how far is the point from mean.

IQR-interquartile range is the difference between the 3rd quartile(75th Percentile ) and the 2nd quartile (25th percentile).

IQR method-1.first find out he IQR by subtracting 3rd Quartile and 2nd quartile

2.Multiply the IQR with 1.5

3.Add it with 3rd quartile. Anything above of the result is a Outlier

4.subtract it with the 2nd quartile . Anything below it is a Outlier

Ans12-In bagging the algorithm learns the weak learners (high bias and variance ) parallelly and combines them using average whereas in boosting the algorithm learns the weak learners sequentially and combines them using a strategy.

Ans13-Adjusted R squared in logistic regression refers to the adjustment made in predicted r squared as increasing the number of variable tend to increase the r squared in Logisitic regression though it does not mean that the model is a good fit .To overcome this problem wile increasing the input variable adjusted r squared is used

Ans 14-Normalization is a scaling technique in which data are shifted and scaled so that it is between 1 and 0.

Standardization-It is a technique in which the values are centered around the mean such that the standard deviation is 1.

Ans15-The algorithm uses the same data for testing that it has used for learning and thus tends to overfit and will not be fruitful when it is provided with real world data. To avoid this problem the training data is split with a parameter k(no. of splits) and the data is provided to the algorithm in a manner that each split/fold is once used for testing and the rest of the cycles for training so that it does not overfit and performs well in real data.

This technique is known as K-fold cross-validation