MACHINE LEARNING

1. C)
2. C)
3. A)
4. A)
5. C)
6. C)
7. C)
8. D)
9. D)
10. A), B) & D)
11. In simple terms, an outlier is an extremely high or extremely low data point relative to the nearest data point and the rest of the neighboring co-existing values in a data graph or dataset we are working with. One common way to find outliers in a dataset is to use the interquartile range. The interquartile range, often abbreviated IQR, is the difference between the 25th percentile (Q1) and the 75th percentile (Q3) in a dataset. It measures the spread of the middle 50% of values. One popular method is to declare an observation to be an outlier if it has a value 1.5 times greater than the IQR or 1.5 times less than the IQR.

12. Bagging and Boosting are two types of Ensemble Learning. These two decrease the variance of a single estimate as they combine several estimates from different models. So, the result may be a model with higher stability. Let's understand these two terms in a glimpse.

a) Bagging: It is a homogeneous weak learners' model that learns from each other independently in parallel and combines them for determining the model average.

b) Boosting: It is also a homogeneous weak learners' model but works differently from Bagging. In this model, learners learn sequentially and adaptively to improve model predictions of a learning algorithm.

13. The adjusted R-squared is a modified version of R-squared that adjusts for the number of predictors in a regression model. It is calculated as: Adjusted $R2 = 1 - [(1-R2)* (n-1)/ (n-k-1)]$ where: R2: The R2 of the model

14. Normalization or Min-Max Scaling is used to transform features to be on a similar scale. The new point is calculated as: X_new = (X - X_min)/(X_max - X_min) and Standardization or Z-Score Normalization is the transformation of features by subtracting from mean and dividing by standard deviation. This is often called as Z-score. X_new = (X - mean)/Std.

15. Cross-validation is a technique in which we train our model using the subset of the data-set and then evaluate using the complementary subset of the data-set. One disadvantage of cross validation is it increases are training time drastically and one advantage of cross validation would be it reduces the chances of overfitting.