**Machine Learning**

1. C
2. C
3. C
4. A
5. C
6. C
7. A
8. B,C
9. A,B,D
10. D
11. An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. In a sense, this definition leaves it up to the analyst (or a consensus process) to decide what will be considered abnormal. Before abnormal observations can be singled out, it is necessary to characterize normal observations.

IQR -Interquartile Range is the difference between the third and the first quartile of a distribution (or the 75th percentile minus the 25th percentile). It is a measure of how wide our distribution is since this range contains half of the points of the dataset. It’s very useful to make an idea of the shape of the distribution. For example, it is the width of the boxes in the boxplot.

1. Differences between Bagging and Boosting. Bagging is the simplest way of combining predictions that belong to the same type while Boosting is a way of combining predictions that belong to the different types. Bagging aims to decrease variance, not bias while Boosting aims to decrease bias, not variance.
2. Adjusted R2 might decrease if a specific effect does not improve the model. Adjusted R squared is calculated by dividing the residual mean square error by the total mean square error (which is the sample variance of the target field). The result is then subtracted from 1. Adjusted R2 is always less than or equal to R2.
3. In statistics, Standardization is the subtraction of the mean and then dividing by its standard deviation. In Algebra, Normalization is the process of dividing of a vector by its length and it transforms your data into a range between 0 and 1.

In Normalization, the change in values is that they are at a standard scale without distorting the differences in the values. Whereas Standardization assumes that the dataset is in Gaussian distribution and measures the variable at different scales, making all the variables equally contribute to the analysis.

1. Cross-validation is a statistical method used to estimate the performance (or accuracy) of machine learning models. It is used to protect against overfitting in a predictive model, particularly in a case where the amount of data may be limited. In cross-validation, you make a fixed number of folds (or partitions) of the data, run the analysis on each fold, and then average the overall error estimate.

Advantages of Cross Validation  
  
1. Reduces Overfitting: In Cross Validation, we split the dataset into multiple folds and train the algorithm on different folds. This prevents our model from overfitting the training dataset. So, in this way, the model attains the generalization capabilities which is a good sign of a robust algorithm.  
  
Note: Chances of overfitting are less if the dataset is large. So, Cross Validation may not be required at all in the situation where we have sufficient data available.  
  
2. Hyperparameter Tuning: Cross Validation helps in finding the optimal value of hyperparameters to increase the efficiency of the algorithm.

Disadvantages of Cross Validation  
  
1. Increases Training Time: Cross Validation drastically increases the training time. Earlier you had to train your model only on one training set, but with Cross Validation you have to train your model on multiple training sets.   
  
For example, if you go with 5 Fold Cross Validation, you need to do 5 rounds of training each on different 4/5 of available data. And this is for only one choice of hyperparameters. If you have multiple choice of parameters, then the training period will shoot too high.  
  
2. Needs Expensive Computation: Cross Validation is computationally very expensive in terms of processing power required.