1. C
2. D
3. C
4. A
5. C
6. B
7. C
8. B,C
9. A,B,C,D
10. A,D
11. An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. IQR is used to measure variability by dividing a data set into quartiles. The data is sorted in ascending order and split into 4 equal parts. Q1, Q2, Q3 called first, second and third quartiles are the values which separate the 4 equal parts.
    Q1 represents the 25th percentile of the data.
    Q2 represents the 50th percentile of the data
    Q3 represents the 75th percentile of the data.
    IQR is the range between the first and the third quartiles namely Q1 and Q3: IQR = Q3 – Q1. The data points which fall below Q1 – 1.5 IQR or above Q3 + 1.5 IQR are outliers.

12. 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.

    In Bagging each model receives equal weight whereas in Boosting models are weighted according to their performance.

    In Bagging each model is built independently whereas in Boosting new models are influenced by performance of previously built models.

    In Bagging different training data subsets are randomly drawn with replacement from the entire training dataset. In Boosting every new subsets contains the elements that were misclassified by previous models.

Bagging tries to solve over-fitting problem while Boosting tries to reduce bias.

Bagging is extended to Random forest model while Boosting is extended to Gradient boosting.

13. Adjusted R2 is a corrected goodness-of-fit (model accuracy) measure for linear models. It identifies the percentage of variance in the target field that is explained by the input or inputs.

    R2 tends to optimistically estimate the fit of the linear regression. It always increases as the number of effects are included in the model. Adjusted R2 attempts to correct for this overestimation. 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. A value of 1 indicates a model that perfectly predicts values in the target field. A value that is less than or equal to 0 indicates a model that has no predictive value. In the real world, adjusted R2 lies between these values
    If the classifier is unstable (high variance), then we should apply Bagging. If the classifier is stable and simple (high bias) then we should apply Boosting.

14. Normalization:

    Minimum and maximum value of features are used for scaling
    It is used when features are of different scales
    Scales values between [0, 1] or [-1, 1].
    It is really affected by outliers.
    Scikit-Learn provides a transformer called MinMaxScaler for Normalization.
    It is useful when we don’t know about the distribution
    It is a often called as Scaling Normalization

Standardization:

Mean and standard deviation is used for scaling

It is used when we want to ensure zero mean and unit standard deviation.

It is not bounded to a certain range.

It is much less affected by outliers.

Scikit-Learn provides a transformer called StandardScaler for standardization

It is useful when the feature distribution is Normal or Gaussian.

It is a often called as Z-Score Normalization

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.

    Cross Validation helps in finding the optimal value of hyperparameters to increase the efficiency of the algorithm.

    Cross Validation drastically increases the training time