Worksheet – 4 ML

Q1 To Q7

1. C
2. C
3. A
4. A
5. A
6. B
7. B

Q8 To Q10

8. B, C
9.
10. D

Q11 To Q15

11. Outliers are those data points that are significantly different from the rest of the dataset.
    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 attempts to tackle the over-fitting issue.

    Boosting tries to reduce bias.

13. Adjusted R^2 is a corrected goodness-of-fit (model accuracy) measure for linear   models. R^2 tends to optimistically estimate the fit of the linear regression.

14. Normalization typically means rescales the values into a range of [0,1].

    Standardization typically means rescales data to have a mean of 0 and a standard deviation of 1 (unit variance).

15. Cross-validation is a statistical method used to estimate the performance (or accuracy) of machine learning models.

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

    Increases Training Time: Cross Validation drastically increases the training time.