

MACHINE LEARNING – 4

- 1.C
- 2.A
- 3.C
- 4.C
- 5.D
- 6.B
- 7.A
- 8.A,C
- 9.A,D
- 10. A,B,D

11. What are outliers? Explain the Inter Quartile Range (IQR) method for outlier detection.

An outlier is an observation that lies at an abnormal distance from other values in a random sample from a population. It is detached from the overall population.

Inter Quartile range is the difference between Q3 and Q1,

Q1-25TH PERCENTILE

Q2-50TH PERCENTILE

Q3-75TH PERCENTILE

12. What is the primary difference between bagging and boosting algorithms?

Bagging specializes in homogenous data set where as Boosting works on Heterogenous data set

13. What is adjusted R^2 in linear regression. How is it calculated?

Adjusted R squared is calculated by dividing the residual mean square error by the total mean square error. Adjusted R^2 is always less than or equal to R^2 .

14. What is the difference between standardisation and normalisation?

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

15. What is cross-validation? Describe one advantage and one disadvantage of using cross-validation.

Cross Validation is a technique which divides the data set into multiple training and testing sets and takes out the mean accuracy.

The advantage of Cross Validation is that it helps in reducing overfitting in the data set where as the disadvantage of Cross Validation is that it increases the training time enormously.