

# Machine learning worksheet 4

Answer no. 1) C

Answer no. 2) C

Answer no. 3) A

Answer no. 4) C

Answer no. 5) A

Answer no. 6) B

Answer no. 7) A

Answer no. 8) B

Answer no. 9) A D

Answer no. 10) A B D

Answer no. 11) a data point on a graph or in a set of results that is very much bigger or smaller than the next nearest data point. Inter quartile range is a most common method to find out outliers in a dataset.

Answer no. 12) In bagging method data is trained parallelly together and in boosting method data is trained in sequence.

Answer no. 13) R square test is used to fit to determine the goodness of fit in regression analysis. Goodness of fit implies how better regression model is fitted to data points.

$$\text{Adjusted } R^2 = 1 - [(1 - R^2) * (n - 1) / (n - k - 1)]$$

Answer no. 14) Normalisation is good if data is not following Gaussian distribution and on the other hand standardisation is good if data is following Gaussian distribution.

Answer no. 15) Cross validation is a technique used to assess a machine learning model and test its performance and accuracy.

Advantage:- It reduces over fitting problem.

Disadvantage :- for inconsistent data it produce different result.

