Anu Sindhuja

MACHINE LEARNING

WORKSHEET – 1

1. C
2. D
3. C
4. D
5. A
6. B
7. C
8. D
9. A,B,C,D
10. A,D
11. Data points that are far away from main group of data which may cause distort statistical results are called outliers. The point at which data falls more than 1.5 times the interquartile range above the third quartile or below the first quartile are called outliers of the data.
12. Bagging tries to solve over-fitting problem.

Boosting tries to reduce bias.

1. Adjusted R2 tells you the percentage of variation explained by only the [independent variables](https://www.statisticshowto.com/independent-variable-definition/) that actually affect the dependent variable. It is calculated based on value of r- squared, number of independent variables, total sample size.
2. Normalization means to scale a variable to a values between 0 and 1, while standardization transforms data to have a mean of zero and a standard deviation of 1.
3. Cross Validation technique deals with overfitting problem in various algorithms.

Advantage: It increases the efficiency of the algorithm by finding the optimal value of hyperparameters.

Disadvantage: It increases the training time.