Machine Learning

1. C
2. C
3. A
4. A
5. c
6. B
7. C
8. D
9. Abcd
10. Abd
11.

An outlier is an extremely high or extremely low data point relative to the nearest data point and the rest of the neighboring co-existing values in a data graph or dataset you're working with.Outliers are extreme values that stand out greatly from the overall pattern of values in a dataset or graph.

The inter quartile range IQR tells us the range where the bulk of the values lie. The inter quartile range is calculated by subtracting the first quartile from the third quartile.

IQR = Q3 - Q1

12.

Bagging is a method of merging the same type of predictions. Boosting is a method of merging different types of predictions. Bagging decreases variance, not bias, and solves over-fitting issues in a model. Boosting decreases bias, not variance.

13.

The adjusted R-squared is a modified version of R-squared that accounts for predictors that are not significant in a regression model. In other words, the adjusted R-squared shows whether adding additional predictors improve a regression model or not.

Calculation of R-squared

Adjusted R Squared = 1 – [((1 – R2) * (n – 1)) / (n – k – 1)]

14.

Normalisation	Standardisation
Scaling is done by the highest and the lowest values.	Scaling is done by mean and standard deviation.
It is applied when the features are of separate scales.	It is applied when we verify zero mean and unit standard deviation.
Scales range from 0 to 1	Not bounded

Affected by outliers	Less affected by outliers
It is applied when we are not sure about the data distribution	It is used when the data is Gaussian or normally distributed
It is also known as Scaling Normalization	It is also known as Z-Score

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.

Advantage of Cross Validation

1. Reduces Over fitting: In Cross Validation, we split the datasets into multiple folds and train the algorithm on different folds. This prevents our model from over fitting the training datasets. So, in this way, the model attains the generalization capabilities which is a good sign of a robust algorithm.

Disadvantage of Cross Validation

1. Increases Training Time: Cross Validation drastically increases the training time. Earlier you had to train your model only on one training set, but with Cross Validation you have to train your model on multiple training sets.