ANSWERSHEET

Machine Learning

QUESTION NO.	ANSWERS
1	C
2	C
3	A
4	A
5	A
6	D
7	C
8	B & C
9	B&D
10	A&D
11	Given below
12	Given below
13	Given below
14	Given below
15	Given below

ANSWER: 11 Outliers are those data points that are significantly different from the rest of the dataset. Inter quartile range is the difference between the first and third quartiles (Q_1 and Q_3). The 'middle half' of the data is between the first and third quartile. The first quartile is the value in the data that separates the bottom 25% of values from the top 75%.The third quartile is the value in the data that separates the bottom 75% of the values from the top 25%. IQR = Q3 – Q1

ANSWER: 12 Bagging is the simplest way of combining predictions that belong to the same type while Boosting is a way of combining predictions that belong to the different types. Bagging aims to decrease variance, not bias while Boosting aims to decrease bias, not variance.

ANSWER: 13 Adjusted R^2 is a corrected goodness-of-fit (model accuracy) measure for linear models. Adjusted R squared is calculated by dividing the residual mean square error by the total mean square error.

ANSWER: 14 Normalization is to transform features to be on a similar scale. Standardization is a process in which the data is restructured in a uniform format.

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.
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

ANSWER: 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. An advantage of using this method is that we make use of all data points and hence it is low bias. The major drawback of this method is that it leads to higher variation in the testing model as we are testing against one data point.