

1. The value of correlation coefficient will always be?
Ans: c. Between -1 and 1.
2. Which of the following cannot be used for dimensionality reduction?
Ans: B. PCA.
3. Which of the following is not a kernel in Support Vector Machines?
Ans: B. Radial Basis Function.
4. Amongst the following, which one is least suitable for a dataset having non-linear decision boundaries?
Ans: A. Logistic Regression.
5. In a Linear Regression problem, 'X' is independent variable and 'Y' is dependent variable, where 'X' represents weight in pounds. If you convert the unit of 'X' to kilograms, then new coefficient of 'X' will be?
Ans: B. same as Old Coefficient of X.
6. As we increase the number of estimators in ADABOOST Classifier, what happens to the accuracy of the model?
Ans: B Increases.
7. Which of the following is not an advantage of using random forest instead of decision trees?
Ans: B. Random Forests explains more variance in data than decision trees
8. Which of the following are correct about Principal Components?
Ans: B) Principal Components are calculated using unsupervised learning techniques
C) Principal Components are linear combinations of Linear Variables.
9. Which of the following are applications of clustering?
Ans: C) Identifying spam or ham emails
D) Identifying different segments of disease based on BMI, blood pressure, cholesterol, blood sugar levels
10. Which of the following is(are) hyper parameters of a decision tree?
Ans: A. Max_depth
D Min_Samples_leaf
11. What are outliers? Explain the Inter Quartile Range (IQR) method for outlier detection.
Ans: outliers are the data points which is located more than the average point of any datasets, n is the no of values in the given set, 25% of the values will be located in $q1 = (\text{data.quantile}(0.25))$, 50 % of the values will be located in $q2$ which is $(\text{data.quantile}(0.50))$ and 75% of the values will be present in $q3$ which is $(\text{data.quantile}(0.75))$
IQR, which is also called as Inter Quartile Range will be detected by $(Q3 - Q1)$
 $IQR = (Q3 - q1)$.
If the outlier point is lying above the $q3$ point, then we can use $q3 + (1.5 * IQR)$
If the outlier Point is lying below the $q1$ point then we can find it by using $q1 - (1.5 * IQR)$

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

Difference Between Bagging and Boosting

Bagging

This method combines predictions that belong to the same type.

Bagging decreases variance.

Base classifiers are trained parallelly.

Boosting

This method combines predictions that belong to the different types.

Boosting decreases bias.

Base classifiers are trained sequentially.

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

Ans: Adjusted R-squared can provide a more precise view of that correlation by also taking into account how many independent variables are added to a particular model against which the Stock index is measured. This is done because such additions of independent variables usually increase the reliability of that model—meaning, for investors, the correlation with the index.

Adjusted $R^2 = 1 - ((1 - R^2)(N - 1) / (N - P - 1))$.

R^2 = r^2 score, N = Number of Samples, P = Number of Independent Variables

14. What is the difference between standardisation and normalisation?

Ans: Standardisation:

Standardization is another scaling method where the values are centered around mean with a unit standard deviation. It means if we will calculate mean and standard deviation of standard scores it will be 0 and 1 respectively.

The formula for standardized values:

$$Z = \frac{x - \mu}{\sigma}$$

Normalisation:

It is a scaling technique method in which data points are shifted and rescaled so that they end up in a range of 0 to 1. It is also known as **min-max scaling**

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

Ans: Cross-Validation is **a statistical method of evaluating and comparing learning algorithms by dividing data into two segments**: one used to learn or train a model and the other used to validate the model.

Advantages of Cross Validation:

1. reduces overfitting
2. hyperparameter tuning can be used to get more accuracy.

Disadvantages of cross validation:

1. Increases Training Time
2. Uses lot of memory and expensive computation.

•