

**In Q1 to Q7, only one option is correct, Choose the correct option:**

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?

3. Which of the following is not a kernel in Support Vector Machines?

4. Amongst the following, which one is least suitable for a dataset having non-linear decision boundaries?

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?

6. As we increase the number of estimators in ADABOOST Classifier, what happens to the accuracy of the model?

7. Which of the following is not an advantage of using random forest instead of decision trees?

**In Q8 to Q10, more than one options are correct, Choose all the correct options:**

8. Which of the following are correct about Principal Components?

Ans :- B & C

9. Which of the following are applications of clustering?

Ans :- B & C

10. Which of the following is(are) hyper parameters of a decision tree?

Ans :- A & D

max\_depth and min\_sample\_leaf

11. What are outliers? Explain the Inter Quartile Range (IQR) method for outlier detection.

Ans :-

In a data collection, outliers are stragglers, which means they

are extremely high **or** extremely low values.

IQR **is** used to measure variability by dividing a data set into quartiles.

IQR **is** the range between the first **and** the third quartiles namely Q1 **and** Q3:

$$IQR = Q3 - Q1.$$

The data points which fall below  $Q1 - 1.5 \text{ IQR}$  **or** above  $Q3 + 1.5 \text{ IQR}$  are outliers.

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

Ans :-

- 1) Bagging attempts to tackle the over-fitting issue. & Boosting tries to reduce bias.
- 2) Every model receives an equal weight. & Models are weighted by their performance.
- 3) Objective to decrease variance, **not** bias. & Objective to decrease bias, **not** variance.

13. What **is** adjusted R2 **in** linear regression. How **is** it calculated?

Ans :-

Adjusted R2 **is** a corrected goodness-of-fit (model accuracy) measure **for** linear models.

It identifies the percentage of variance **in** the target field that **is** explained by the input **or** inputs.

It **is** calculated by

$$R^2 = \{(1 / N) * \sum [(x_i - \bar{x}) * (Y_i - \bar{y})] / (\sigma_x * \sigma_y)\}^2$$

14. What **is** the difference between standardisation **and** normalisation?

Ans :-

- 1) Scales range from 0 to 1 & **not** bounded
- 2) Affected by outliers & less affected by outliers
- 3) It **is** also known as Scaling Normalization & it **is** also known as Z-Score

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

Ans :-

Cross-validation **is** a procedure to evaluate the performance of learning models.

Advantage :-

Hyperparameter Tuning: Cross Validation helps **in** finding the optimal value of hyperparameters to increase the efficiency of the algorithm.

Disadvantage :-

Needs Expensive Computation: Cross Validation **is**  
computationally very expensive **in**  
terms of processing power required.