MACHINE LEARNING

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

1. The value of correlation coefficient will always be:
   A) between 0 and 1 B) greater than -1
   C) between -1 and 1 D) between 0 and -1
   Answer :- C
2. Which of the following cannot be used for dimensionality reduction?
   A) Lasso Regularisation **B) PCA**
   C) Recursive feature elimination D) Ridge Regularisation
   Answer :- B
3. Which of the following is not a kernel in Support Vector Machines?
   A) **linear** B) Radial Basis Function
   C) hyperplane D) polynomial
   Answer :- A

4. Amongst the following, which one is least suitable for a dataset having non-linear decision boundaries?
   A) **Logistic Regression** B) Naïve Bayes Classifier
   C) Decision Tree Classifier D) Support Vector Classifier
   Answer :- A

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?
   (1 kilogram = 2.205 pounds)
   A) 2.205 × old coefficient of 'X' B) same as old coefficient of 'X'
   C) old coefficient of 'X' ÷ 2.205 D) Cannot be determined
   Answer :- A

6. As we increase the number of estimators in ADABOOST Classifier, what happens to the accuracy of the model?
   A) remains same **B) increases**
   C) decreases D) none of the above
   Answer :- B

7. Which of the following is not an advantage of using random forest instead of decision trees?
   A) Random Forests reduce overfitting
   B) Random Forests explains more variance in data then decision trees
   C) Random Forests are easy to interpret
   D) Random Forests provide a reliable feature importance estimate
   Answer: - A

MACHINE LEARNING

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?
   A) Principal Components are calculated using supervised learning techniques
   B) Principal Components are calculated using unsupervised learning techniques
   C) Principal Components are linear combinations of Linear Variables.
   D) All of the above
   Answer :- D
9. Which of the following are applications of clustering?
   A) Identifying developed, developing and under-developed countries on the basis of factors like GDP, poverty index, employment rate, population and living index
   B) Identifying loan defaulters in a bank on the basis of previous years' data of loan accounts.
   C) Identifying spam or ham emails
   D) Identifying different segments of disease based on BMI, blood pressure, cholesterol, blood sugar levels.
   Answer :- D

10. Which of the following is(are) hyper parameters of a decision tree?
    A) max_depth B) max_features
    C) n_estimators D) min_samples_leaf

    Answer :- D

MACHINE LEARNING

Q10 to Q15 are subjective answer type questions, Answer them briefly.

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

Ans:- Definition of outliers. An outlier is **an observation that lies an abnormal distance from other values in a random sample from a population**.

Once we calculate it, we can use IQR to identify the outliers. We label a point as an outlier if it satisfies one of the following conditions:

- It's greater than 75th percentile + 1.5 IQR
- It's less than 25th percentile – 1.5 IQR

Applying this simple formula, we can easily detect the outliers of our distribution. Boxplot uses the same method to plot the outliers as points outside the whiskers.

The reasons behind that 1.5 coefficient rely upon the normal distribution, but the general idea is to calculate outliers without using some measure that could be affected by them. That's why using, for example, the standard deviation, could lead us to poor results. Quartiles and percentiles are based on counts, so they are less vulnerable to the presence of outliers.

The idea is that if a point is too far from the 75th percentile (or from the 25th percentile), it's a "strange" point that can be labeled as an outlier. The order of magnitude of such a distance is the IQR itself.

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

Answer:- Differences Between Bagging and Boosting

S.NO	Bagging	Boosting
1.	The simplest way of combining predictions that belong to the same type.	A way of combining predictions that belong to the different types.
2.	Aim to decrease variance, not bias.	Aim to decrease bias, not variance.
3.	Each model receives equal weight.	Models are weighted according to their performance.
4.	Each model is built independently.	New models are influenced by the performance of previously built models.
5.	Different training data subsets are selected using row sampling with replacement and random sampling methods from the entire training dataset.	Every new subset contains the elements that were misclassified by previous models.
6.	Bagging tries to solve the over-fitting problem.	Boosting tries to reduce bias.
7.	If the classifier is unstable (high variance), then apply bagging.	If the classifier is stable and simple (high bias) the apply boosting.
8.	In this base classifiers are trained parallelly.	In this base classifiers are trained sequentially.
9	Example: The Random forest model uses Bagging.	Example: The AdaBoost uses Boosting techniques

MACHINE LEARNING

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

Ans:- Adjusted R-squared is **a modified version of R-squared that has been adjusted for the number of predictors in the model**. The adjusted R-squared increases when the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected.

R-squared, often written R^2, is the proportion of the variance in the response variable that can be explained by the predictor variables in a linear regression model.

The value for R-squared can range from 0 to 1. A value of 0 indicates that the response variable cannot be explained by the predictor variable at all while a value of 1 indicates that the response variable can be perfectly explained without error by the predictor variables.

The **adjusted R-squared** is a modified version of R-squared that adjusts for the number of predictors in a regression model. It is calculated as:

Adjusted $R^2 = 1 – [(1-R^2)*(n-1)/(n-k-1)]$

where:

- **R^2**: The R^2 of the model
- **n**: The number of observations
- **k**: The number of predictor variables

Because R^2 always increases as you add more predictors to a model, adjusted R^2 can serve as a metric that tells you how useful a model is, *adjusted for the number of predictors in a model*.

14. What is the difference between standardisation and normalisation?

Answer:- **Feature scaling** is one of the most important data preprocessing step in machine learning. Algorithms that compute the distance between the features are biased towards numerically larger values if the data is not scaled.
Tree-based algorithms are fairly insensitive to the scale of the features. Also, feature scaling helps machine learning, and deep learning algorithms train and converge faster.

There are some feature scaling techniques such as Normalization and Standardization that are the most popular and at the same time, the most confusing ones.

Difference between Normalization and Standardization

S.NO.	Normalization	Standardization
1.	Minimum and maximum value of features are used for scaling	Mean and standard deviation is used for scaling.
2.	It is used when features are of different scales.	It is used when we want to ensure zero mean and unit standard deviation.

MACHINE LEARNING

S.NO.	Normalization	Standardization
3.	Scales values between [0, 1] or [-1, 1].	It is not bounded to a certain range.
4.	It is really affected by outliers.	It is much less affected by outliers.
5.	Scikit-Learn provides a transformer called `MinMaxScaler` for Normalization.	Scikit-Learn provides a transformer called `StandardScaler` for standardization.
6.	This transformation squishes the n-dimensional data into an n-dimensional unit hypercube.	It translates the data to the mean vector of original data to the origin and squishes or expands.
7.	It is useful when we don't know about the distribution	It is useful when the feature distribution is Normal or Gaussian.
8.	It is a often called as Scaling Normalization	It is a often called as Z-Score Normalization.

15. What is cross-validation? Describe one advantage and one disadvantage of using cross-validation.
Answer:- Cross validation is **a technique for assessing how the statistical analysis generalises to an independent data set**.It is a technique for evaluating machine learning models by training several models on subsets of the available input data and evaluating them on the complementary subset of the data.
Advantage:-
An advantage of using this method is that we make use of all data points and hence it is low bias.
Disadvantage:-
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. If the data point is an outlier it can lead to higher variation. Another drawback is it takes a lot of execution time as it iterates over 'the number of data points' times.