FLIP ROBO ASSIGNMENT - 4 **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

Ans : C)between -1 and 1

2. Which of the following cannot be used for dimensionality reduction?
   A) Lasso Regularization B) PCA
   C) Recursive feature elimination D) Ridge Regularization

Ans : A)Lasso Regularization , D)Ridge Regularization

3. Which of the following is not a kernel in Support Vector Machines?
   A) linear B) Radial Basis Function
   C) hyperplane D) polynomial

Ans :C)hyperplane

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

Ans :Decision Tree Classifier

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

Ans : A)2.205 x old coefficient of ‘X

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

Ans : B)increases

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

Ans : C)Random Forests are easy to interpret

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

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?

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.

Ans :A),C) and 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

Ans : A),B),D)

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

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

Ans :An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. In a sense, this definition leaves it up to the analyst (or a consensus process) to decide what will be considered abnormal. Before abnormal observations can be singled out, it is necessary to characterize normal observations.

IQR(Interquartile range):

IQR is used to measure variability by dividing a data set into quartiles. The data is sorted in ascending order and split into 4 equal parts. Q1, Q2, Q3 called first, second and third quartiles are the values which separate the 4 equal parts.

Q1 represents the 25th percentile of the data.

Q2 represents the 50th percentile of the data.

Q3 represents the 75th percentile of the data.

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

Ans : Bagging

1)The simplest way of combining predictions that belong to the same type.

2. Aim to decrease variance, not bias.

3. Each model receives equal weight.

4. Each model is built independently. .

5. Different training data subsets are selected using row sampling with replacement and random sampling methods from the entire training dataset.

6. Bagging tries to solve the over-fitting problem. .

7. If the classifier is unstable (high variance), then apply bagging.

8. In this base classifiers are trained parallelly.

9 Example: The Random forest model uses Bagging.

Boosting

1)A way of combining predictions that belong to the different types.

2.Aim to decrease bias, not variance.

3.Models are weighted according to their performance.

4.New models are influenced by the performance of previously built models.

5.Every new subset contains the elements that were misclassified by previous models.

6.Boosting tries to reduce bias.

7.If the classifier is stable and simple (high bias) the apply boosting.

8. In this base classifiers are trained sequentially.

9)Example: The AdaBoost uses Boosting techniques

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

Ans : R2 shows how well terms (data points) fit a curve or line. Adjusted R2 also indicates how well terms fit a curve or line, but adjusts for the number of terms in a model. If you add more and more useless variables to a model, adjusted r-squared will decrease. If you add more useful variables, adjusted r-squared will increase.Adjusted R2 will always be less than or equal to R2.

You only need R2 when working with samples. In other words, R2 isn't necessary when you have data from an entire population.
The formula is:

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right]$$

where:

- N is the number of points in your data sample.
- K is the number of independent regressors, i.e. the number of variables in your model, excluding the constant.

14. What is the difference between standardisation and normalisation?

Ans :

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.
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	It is a often called as Z-Score

Scaling Normalization Normalization.

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.

Advantage of Cross Validation

1. Reduces Overfitting: In Cross Validation, we split the dataset into multiple folds and train the algorithm on different folds. This prevents our model from overfitting the training dataset. 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.