**NAME OF THE PROJECT**

**Machine learning  
Submitted by:**

**SAUNAK MUKHERJEE**



1. **The value of correlation coefficient will always be:**

C) between -1 and 1

1. **Which of the following cannot be used for dimensionality reduction?**

B) PCA

1. **Which of the following is not a kernel in Support Vector Machines?**
2. linear
3. **Amongst the following, which one is least suitable for a dataset having non-linear decision boundaries**?

D) Support Vector 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?**

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

**the model?**

1. decreases

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

B) Random Forests explains more variance in data then decision trees

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

B) max\_features

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

D) All of the above

**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

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

Something which is is an observation that lies an abnormal distance from other values in a random sample from a population.

It is the range between the first and the third quartiles namely Q1 and Q3: IQR = Q3 – Q1.

**s12. What is the primary difference between bagging and boosting algorithm?**

Bagging which is the technique for reducing prediction variance by producing additional data for training from a dataset by combining repetitions with combinations to create multi-sets of the original data. Boosting is an iterative strategy for adjusting an observation's weight based on the previous classification.

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

Adjusted R2 that is a corrected goodness-of-fit measure for linear models. It is also check the model accuracy.

R 2 = 1 − sum squared regression (SSR) total sum of squares (SST) , = 1 − ∑ ( y i − y i ^ ) 2 ∑ ( y i − y ¯ ) 2 .

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

Normalization

is a often called as Scaling Normalization and standardisation is a often called as Z-Score Normalization.

Normalization is used when the data doesn't have Gaussian distribution whereas Standardization is used on data having Gaussian distribution.

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

Cross-Validation which is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments

cross-validation Using the rest data-set train the model.