Q1.The value of correlation coefficient will always be: -1 and 1

Q2. Which of the following cannot be used for dimensionality reduction? Ridge Regularisation

Q3. Which of the following is not a kernel in Support Vector Machines? Hyperplane

Q4. Amongst the following, which one is least suitable for a dataset having non-linear decision boundaries?Naïve Bayes Classifier

Q5. . 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? same as old coefficient of ‘X’

Q6. As we increase the number of estimators in ADABOOST Classifier, what happens to the accuracy of the model? Increases

Q7. Which of the following is not an advantage of using random forest instead of decision trees?Random Forests provide a reliable feature importance estimate

Q8. . Which of the following are correct about Principal Components? Principal Components are calculated using unsupervised learning techniques, Principal Components are linear combinations of Linear Variables.

Q9. Which of the following are applications of clustering? Identifying different segments of disease based on BMI, blood pressure, cholesterol, blood sugar levels, 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

Q10. Which of the following is(are) hyper parameters of a decision tree? max\_depth , min\_samples\_leaf

Q11.What are outliers? Explain the Inter Quartile Range(IQR) method for outlier detection?

Outliers are the values which are lies outside the range .Inter Quartile Rule is IQR= Q3 – Q1

A commonly used rule says that a data point is an outlier if it is more than 1.5 .

Q12. . What is the primary difference between bagging and boosting algorithms?

Bagging is a way to decrease the variance in the prediction by generating additional data for training from dataset using combinations with repetitions to produce multi-sets of the original data. Boosting is an iterative technique which adjusts the weight of an observation based on the last classification.

Q13. . What is adjusted R2 in logistic regression. How is it calculated?

The adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. Adjusted R-squared value can be calculated based on value of r-squared, number of independent variables (predictors), total sample size. Every time you add a independent variable to a model, the R-squared increases, even if the independent variable is insignificant.

Q14. What is the difference between standardisation and normalisation?

The terms normalization and standardization are sometimes used interchangeably, but they usually refer to different things. Normalization usually means to scale a variable to have a values between 0 and 1, while standardization transforms data to have a mean of zero and a standard deviation of 1.

Q15. What is cross-validation? Describe one advantage and one disadvantage of using cross-validation.

Cross validation is a model evaluation method that is better than residuals. The problem with residual evaluations is that they do not give an indication of how well the learner will do when it is asked to make new predictions for data it has not already seen. One way to overcome this problem is to not use the entire data set when training a learner. Some of the data is removed before training begins. Then when training is done, the data that was removed can be used to test the performance of the learned model on ``new'' data. This is the basic idea for a whole class of model evaluation methods called *cross validation*.

Advantage:-  the proportion of the validation or training split is not dependent on the number of folds (K-fold test).

Disadvantage:- There are chances that you might miss out some observations whereas you might select some observations more than once.