MACHINE LEARNING-4

Answers of following Questions-

Ans 1-(C) between -1 and 1

Ans 2-(B) PCA

Ans 3-(A) linear

Ans 4-(D) Support Vector Classifier

Ans 5-(A) 2.205 × old coefficient of 'X'

Ans 6-(B) increases

Ans 7-(B) Random Forests explains more variance in data then decision trees

Ans 8-(A) Principal Components are calculated using supervised learning techniques

Ans 9-(C) Identifying spam or ham emails

Ans 10-(D) min_samples_leaf

Ans 11-An outlier is an object that deviates significantly from the rest of the objects. They can be caused by measurement or execution error

Inter Quartile Range

Find the first quartile, Q1.

Find the third quartile, Q3.

Calculate the IQR. IQR = Q3-Q1.

Define the normal data range with lower limit as Q1–1.5*IQR and upper limit as Q3+1.5*IQR.

Any data point outside this range is considered as outlier and should be removed for further analysis

Ans 12-Bagging is a method of merging the same type of predictions. Boosting is a method of merging different types of predictions. Bagging decreases variance, not bias, and solves over-fitting issues in a model. Boosting decreases bias, not variance

Ans 13-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. It never declines

Ans 14-Normalisation, the change in values is that they are at a standard scale without distorting the differences in the values. Whereas, Standardisation assumes that the dataset is in Gaussian distribution and measures the variable at different scales, making all the variables equally contribute to the analysis

Ans 15-Cross-validation, sometimes called rotation estimation or out-of-sample testing, is any of various similar model validation techniques for assessing how the results of a statistical analysis will generalize to an independent data set

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

2. 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