TASK 2

1. (C) -1 to 1

2. (D) Ridge Regularisation

3. (C) Hyperplane

4. (D) Support Vector Classifier

5. (C)

6. (B) If we increase the number of estimators in ADABOOST Classifier then the the accuracy of the model "increases".

7. (A)

8. (D) B) Principal Components are calculated using unsupervised learning techniques.

C) Principal Components are linear combinations of Linear Variables.

9. A,B,C and D

10.A,B,C

11. OUTLIER : An outlier is a data point that differs significantly from other observations.

INTERQUARTILE RANGE : The interquartile range (IQR) is a measure of variability, based on dividing a data set into quartiles.

Quartiles divide a rank-ordered data set into four equal parts.

The values that divide each part are called the first, second, and third quartiles; and they are denoted by Q1, Q2, and Q3, respectively.

12. "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.

13. The 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 only if 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 by chance.

Adjusted R-squared value can be calculated based on value of r-squared, number of independent variables (predictors), total sample size.

14. STANDARDISATION : Standardization refers to shifting the distribution of each attribute to have a mean of zero and a standard deviation of one (unit variance).

It is useful to standardize attributes for a model that relies on the distribution of attributes such as Gaussian processes

NORMALISATION : Normalization refers to rescaling real-valued numeric attributes into a 0 to 1 range.

Normalization makes the features more consistent with each other, which allows the model to predict outputs more accurately.

15. CROSS VALIDATION : Cross-validation is a technique in which we train our model using the subset of the data-set and then evaluate using the complementary subset of the data-set.

ADVANTAGE : Cross Validation is a very useful technique to overcome the problem of overfitting.

DISADVANTAGE : Cross validation process can become a lengthy one.