***Support Vector Machine Model and RBF kernel***

- First we begin by uploading our data which consists of the grades of the student thorough the semester and the total grade and the class letter G for “Good” >60 and W for “Weak” <=59

- Dropping unnecessary column “Student ID”

- Initializing X and Y at 20% of the data **(quiz1, assignment1)**

- Finding that the data is imbalanced

- Solving the Imbalanced data using “smote” technique

SMOTE ALGORITHM

SMOTE is an algorithm that performs data augmentation by creating **synthetic data points** based on the original data points. SMOTE can be seen as an advanced version of oversampling, or as a specific algorithm for data augmentation. The advantage of SMOTE is that you are **not generating duplicates**, but rather creating synthetic data points that are **slightly different** from the original data points. You draw a random sample from the minority class.

- checking that the data is now balanced **{'G': 465, 'W': 465}**

- applying feature scaling

Feature Scaling

Feature Scaling is a technique to standardize the independent features present in the data in a fixed range. It is performed during the data pre-processing to handle highly varying magnitudes or values or units. If feature scaling is not done, then a machine learning algorithm tends to weigh greater values, higher and consider smaller values as the lower values, regardless of the unit of the values.

- Fitting Data into SVM-RBF model

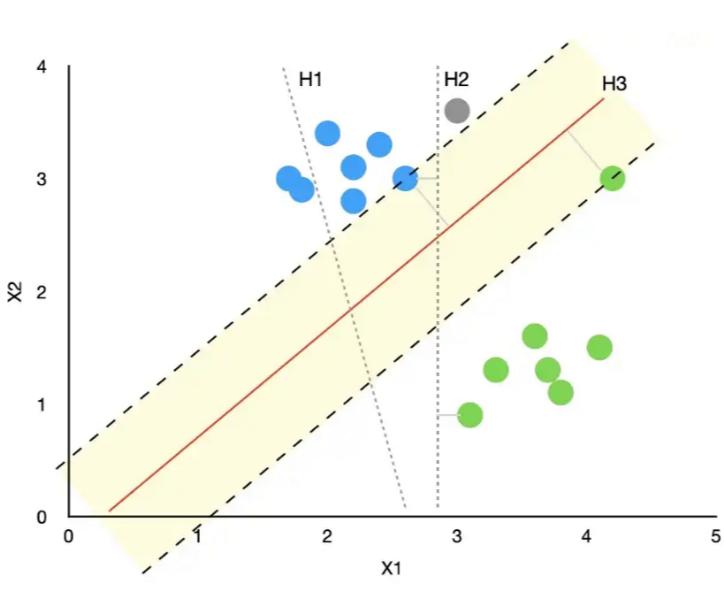
SVM Model

Classification issues are dealt with most frequently by Support Vector Machines (SVMs), which are a subset of supervised machine learning.

The SVM algorithm looks for a hyperplane that divides these two classes by the greatest margin. A hard-margin can be used if classes are completely linearly separable. Otherwise, a soft-margin is required.

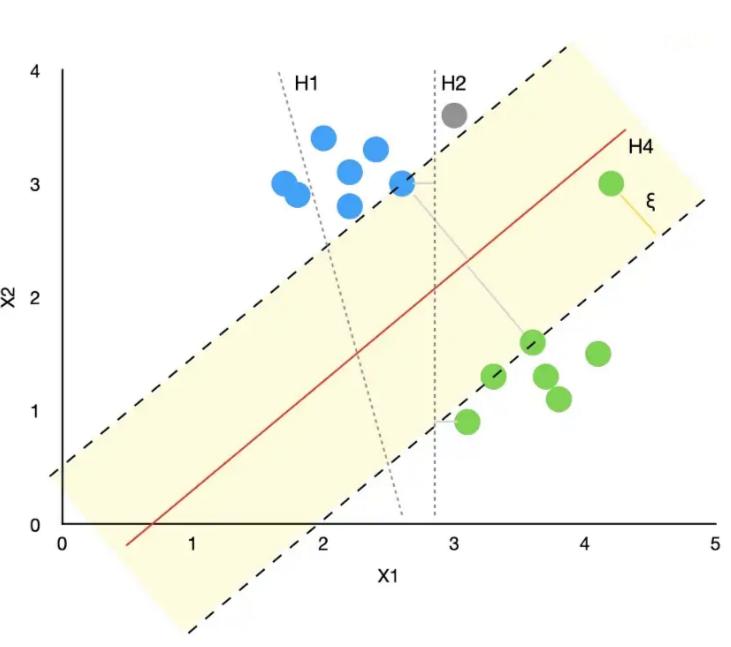
Hard Margin

The “H3” hyperplane separates the two classes correctly and with the highest possible margin



Soft Margin

We can see that the “H4” hyperplane treats the green point inside the margin as an outlier. Hence, the support vectors (data points that are closer to the hyper plane and influence the position and orientation of the hyperplane) are the two green points closer to the main group of green points. This allows a larger margin to exist, increasing the model’s robustness.



RBF Kernel

A kernel is a function that takes the original non-linear problem and transforms it into a linear one within the higher-dimensional space.

C = [0.25,0.5,1] C- It is a hypermeter in SVM to control error

gamma = range (0.3,0.8) Gamma is a hyperparameter which we have to set before training model. Gamma decides that how much curvature we want in a decision boundary.

Gamma high means more curvature.

Gamma low means less curvature.

- Displaying the accuracy: Accuracy of this model at 20% is: 86.26609442060087 %

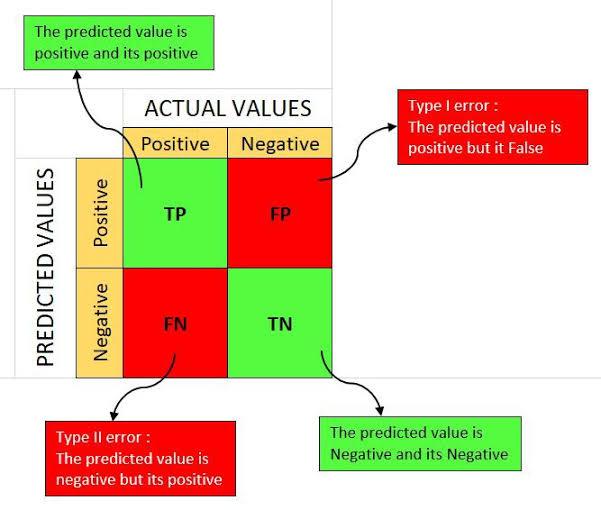
- Displaying Confusion matrix

[[ 93 21]

[ 11 108]]

Confusion matrix

Presents a table layout of the different outcomes of the prediction and results of a classification problem and helps visualize its outcomes.



- Repeat all the steps again with 50% of the data **(quiz1, assignment1, midterm, assignemt2)**

- Displaying the accuracy: Accuracy of this model at 50% is: 96.56652360515021 %

- Displaying Confusion matrix

[[106 8]

[ 0 119]]