# AML Assignment-2

## 2.

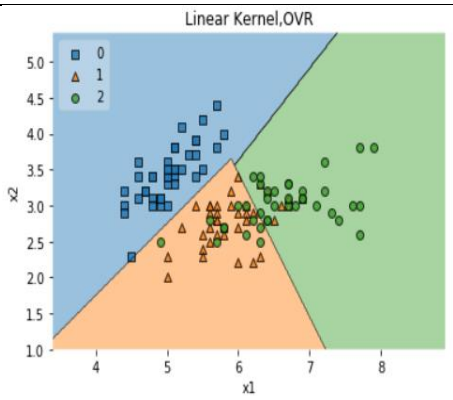
### I, II. Decision boundary:

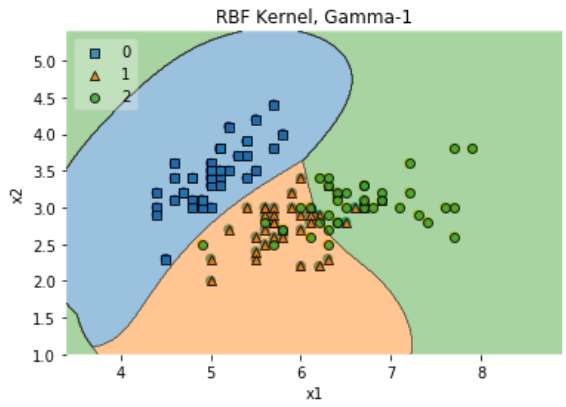
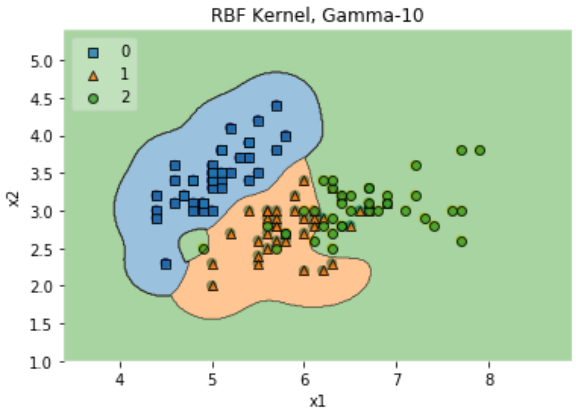
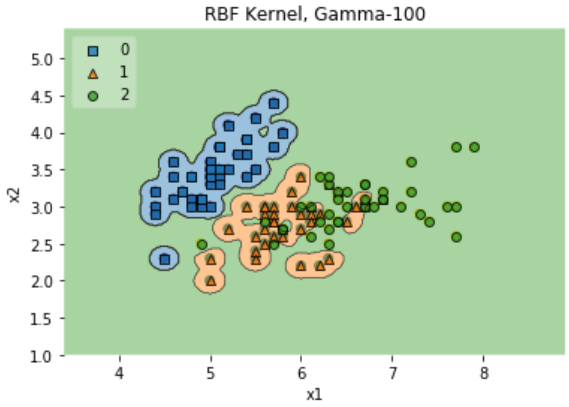
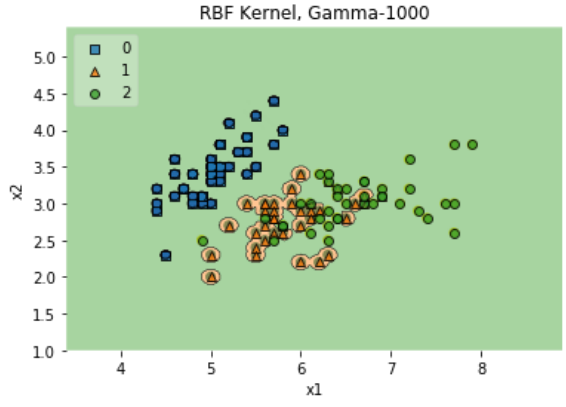
Dataset: Iris dataset

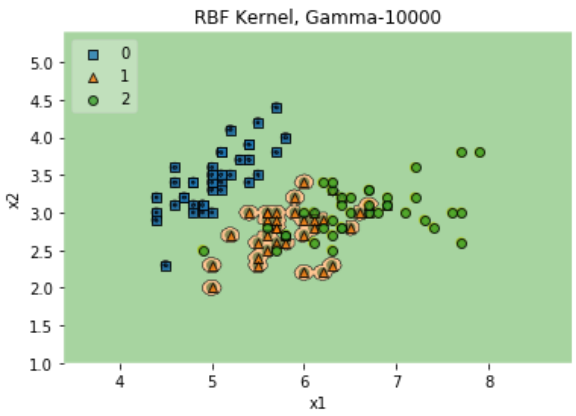
- This dataset has 150 data points and 3 classes (multi-class dataset) and each data point has 4 features.
- We can classify the points by taking any two features, because each data point is different.
- Given class labels as Iris-setosa:0, Iris-versicolor:1, Iris-virginica:2 in pre-processing step.
- I splitted the dataset into train and test sets.

Training:

- Trained the dataset using 'Linear' kernel and 'Gaussian (RBF)' kernel with γ= {1, 10, 100, 1000, 10000} separately and plotted the decision boundaries.
- As SVM is Binary classification technique, and our dataset is multi-class dataset. So, I implemented algorithm by using OVR (One-Vs.-Rest) and OVO (One-Vs.-One) methods.
- OVR method: For one-class set to '1' and remaining classes as '0'. Do the procedure for remaining classes. If there are 'n' classes, then 'n' classifiers will be built. Here we have #classes=3, so built 3 classifiers.
    - For testing, we will take max probability decision function rule.
- OVO method: We will take two classes each time and do the training. If there are 'n' classes, then 'n*(n-1)/2' classifiers will be built. Here we have #classes=3, so built 3*2/2=3 classifiers.
    - For testing, we will predict based on maximum voting.ad

| Model | Decision Boundary plot | Metric |
|---|---|---|
| Linear Kernel |   | Training Accuracy=<br>Test Accuracy=<br>Good |

| | | |
|---|---|---|
| RBF Kernel Gamma=1 |  | Train Accuracy: 0.8416666666666667<br>Test Accuracy: 0.7666666666666667<br>Good |
| RBF Kernel Gamma=10 |  | Train Accuracy: 0.875<br>Test Accuracy: 0.7333333333333333<br>Good |
| RBF Kernel Gamma=100 |  | Train Accuracy: 0.9333333333333333<br>Test Accuracy: 0.6666666666666666<br>Overfits |
| RBF Kernel Gamma=1000 |  | Train Accuracy: 0.9416666666666667<br>Test Accuracy: 0.5333333333333333<br>Overfits |

| | | |
|---|---|---|
| RBF Kernel Gamma=10000 |  | Train Accuracy: 0.9416666666666667<br>Test Accuracy: 0.5333333333333333<br>Overfits |

## iii. OVR Linear Kernel KFold:

-Done the OVR on linear kernel and plotted the decision boundaries for each fold(taken KFold=10).

-Analysed the change of accuracies for each fold.

(- See pynotebook for results and plots)

### For RBF Kernel:

-For each fold, done the training and calculated the accuracies for training, validation and for testing.

-Analysed the behaviour of Gamma value according to each fold and plotted decision boundaries for them (See plots folder and code notebook).

# 3. Online SVM:

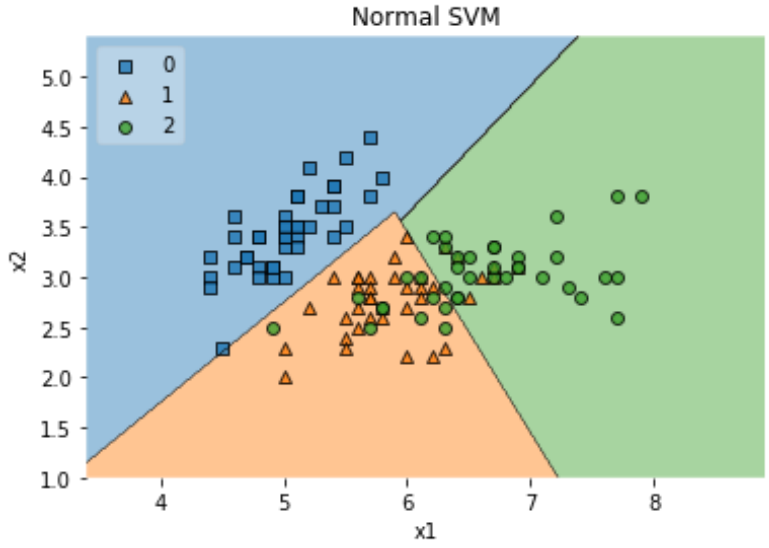
Online SVM is a concept which accommodates new training data for the purpose of better classification. The steps involved are:

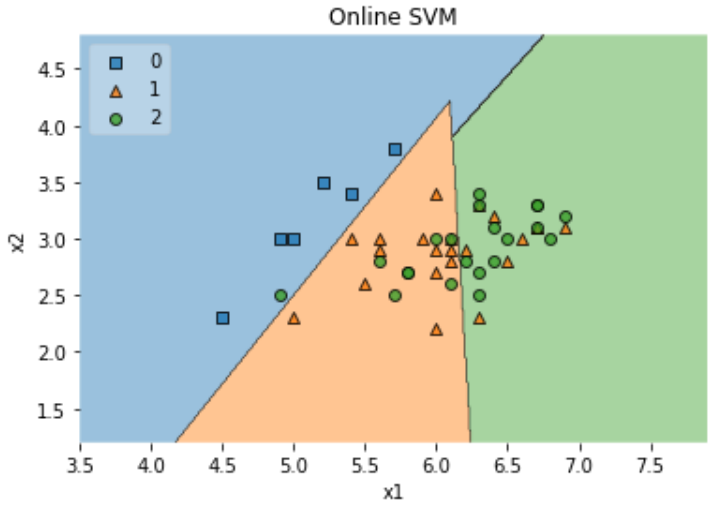
Algorithm used:

- Initially trained on some subset of dataset.
- Now, each time new data point and try to predict it, if it is misclassified, then the point is added to the set of support vectors and the model will be trained again along with that new data point .
- By retraining it, the support vectors may be increased.
- SVM's as the addition of a new data point can change the decision boundary thereby resulting in over fitting/under-fitting.
- If it is classified correctly then ignore that data point.
- the accuracy increased on each step of the online SVM training are there are less no. of data points and learning every data point will translate to a 100% accurate model.
- Experimented on Linear and RBF kernels.
- In both of them, by using online svm, it is approximately giving accuracy as like normal svm.

Linear SVM:

TrainAcc: 0.8416666666666667 TestAcc: 0.7333333333333333
TrainAcc: 0.8416666666666667 TestAcc: 0.7






RBF Kernel:

Training Accuracy:0.85, TestAccuracy:0.73

Training Accuracy:0.82, TestAccuracy:0.7