CS 584 – Machine Learning

Assignment 4

Report

SUPPORT VECTOR MACHINES:

1) Generate 2D dataset:

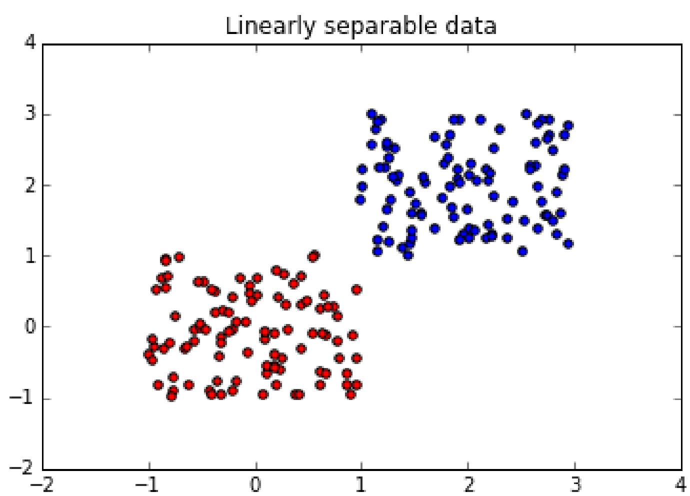
For generating the datasets, random numbers are generated using numpy's in-built function and the distance between the points are adjusted so as to get the data that is linearly separable and non-linearly separable.

For e.g.,

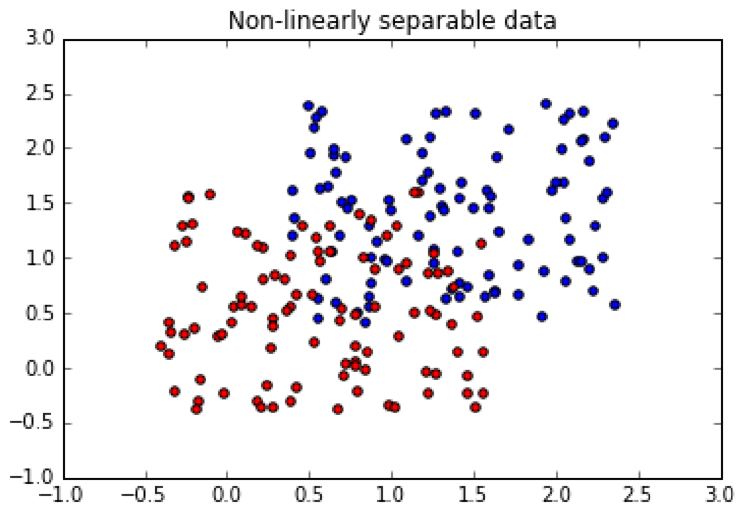
While constructing the dataset if we designate the distance to be '1.0' then we end up in a linearly separable data.

If the distance is '0.5' or less than '0.5', the resulting dataset would be non-linearly separable. Below are the graphs of the generated datasets:

a) Linearly separable data:



b) Non-linearly separable data:



2) Linear SVM with hard margins:

Since the data in both the dimensions are in the same range, there is no need to normalize the data.

<u>Results observed for linearly separable data:</u>

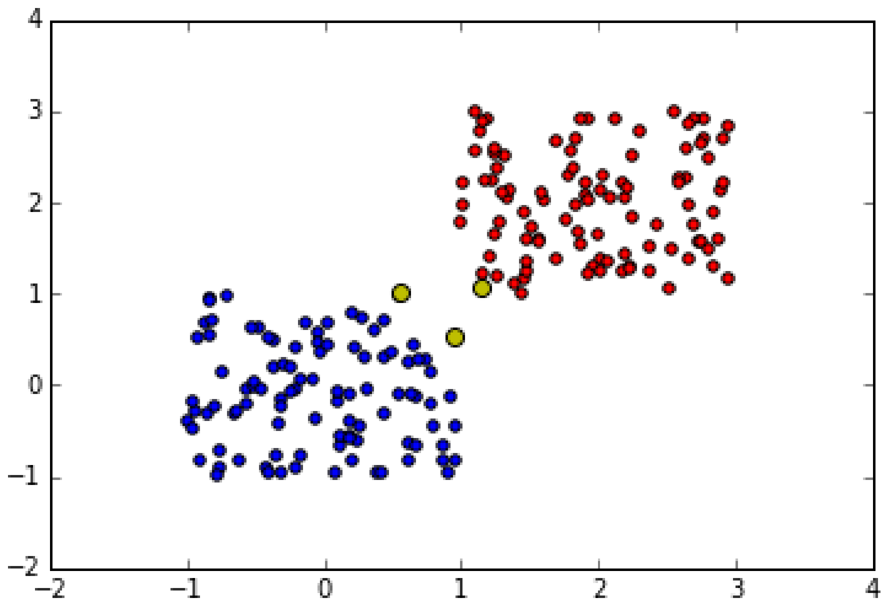
```
pcost       dcost       gap    pres   dres
 0: -2.5464e+01 -4.8317e+01  5e+02  2e+01  2e+00
 1: -3.5263e+01 -2.5704e+01  2e+02  8e+00  7e-01
 2: -8.7654e+01 -5.7413e+01  2e+02  7e+00  6e-01
 3: -1.1652e+02 -6.0320e+01  2e+02  5e+00  5e-01
 4: -3.4238e+01 -1.3506e+01  6e+01  1e+00  1e-01
 5: -7.1555e+00 -9.5265e+00  2e+00  4e-15  2e-14
 6: -8.4551e+00 -8.4917e+00  4e-02  6e-15  1e-14
 7: -8.4811e+00 -8.4815e+00  4e-04  2e-15  1e-14
 8: -8.4814e+00 -8.4814e+00  4e-06  3e-15  1e-14
Optimal solution found.
No of. support vectors found 3
Predicted  -1   1
Truth
-1          9   0
 1          0  11
     pcost       dcost       gap    pres   dres
 0: -2.5303e+01 -4.7868e+01  5e+02  2e+01  2e+00
 1: -3.8535e+01 -2.7387e+01  2e+02  8e+00  7e-01
 2: -8.3125e+01 -5.5377e+01  2e+02  7e+00  6e-01
 3: -1.4178e+02 -7.3327e+01  2e+02  6e+00  5e-01
```

```
 4: -3.4975e+01 -1.4244e+01  8e+01  1e+00  1e-01
 5: -6.8825e+00 -9.6448e+00  3e+00  2e-14  3e-14
 6: -8.4450e+00 -8.4928e+00  5e-02  6e-15  1e-14
 7: -8.4810e+00 -8.4815e+00  5e-04  1e-14  1e-14
 8: -8.4814e+00 -8.4814e+00  5e-06  9e-15  1e-14
Optimal solution found.
No of. support vectors found 3
Predicted  -1   1
Truth
-1         13   0
 1          0   7
     pcost       dcost       gap    pres   dres
 0: -2.4737e+01 -4.7428e+01  5e+02  2e+01  2e+00
 1: -3.4365e+01 -2.6747e+01  2e+02  8e+00  7e-01
 2: -8.4743e+01 -5.8323e+01  2e+02  7e+00  7e-01
 3: -1.3791e+02 -6.5204e+01  2e+02  5e+00  4e-01
 4: -1.5041e+01 -1.1719e+01  3e+01  5e-01  4e-02
 5: -7.3668e+00 -8.9830e+00  2e+00  2e-15  1e-14
 6: -8.4670e+00 -8.4866e+00  2e-02  1e-14  9e-15
 7: -8.4812e+00 -8.4814e+00  2e-04  2e-14  1e-14
 8: -8.4814e+00 -8.4814e+00  2e-06  6e-15  1e-14
Optimal solution found.
No of. support vectors found 3
Predicted  -1   1
Truth
-1         10   0
 1          0  10
     pcost       dcost       gap    pres   dres
 0: -2.4445e+01 -4.6626e+01  5e+02  2e+01  2e+00
 1: -3.7192e+01 -2.7364e+01  2e+02  8e+00  7e-01
 2: -8.8456e+01 -5.9398e+01  2e+02  7e+00  6e-01
 3: -1.3245e+02 -6.5046e+01  2e+02  5e+00  4e-01
 4: -2.1487e+01 -1.2062e+01  5e+01  8e-01  7e-02
 5: -7.2031e+00 -9.2562e+00  2e+00  8e-15  2e-14
 6: -8.4592e+00 -8.4898e+00  3e-02  8e-15  1e-14
 7: -8.4811e+00 -8.4815e+00  3e-04  9e-15  1e-14
 8: -8.4814e+00 -8.4814e+00  3e-06  1e-14  9e-15
Optimal solution found.
No of. support vectors found 4
Predicted  -1   1
Truth
-1         11   0
 1          0   9
     pcost       dcost       gap    pres   dres
 0: -2.5201e+01 -4.7801e+01  5e+02  2e+01  2e+00
 1: -3.6729e+01 -2.6069e+01  2e+02  8e+00  7e-01
 2: -6.4970e+01 -4.2902e+01  2e+02  7e+00  6e-01
 3: -5.2322e+01 -2.0146e+01  1e+02  3e+00  3e-01
 4: -8.9104e+00 -7.0942e+00  1e+01  3e-01  2e-02
 5: -5.7928e+00 -6.0773e+00  3e-01  2e-15  8e-15
```

```
 6: -5.9978e+00 -6.0009e+00  3e-03  2e-15  8e-15
 7: -6.0001e+00 -6.0001e+00  3e-05  4e-15  8e-15
 8: -6.0001e+00 -6.0001e+00  3e-07  9e-16  9e-15
Optimal solution found.
No of. support vectors found 3
Predicted  -1   1
Truth
-1         11   1
 1          0   8
     pcost       dcost       gap    pres   dres
 0: -2.3770e+01 -4.5197e+01  5e+02  2e+01  2e+00
 1: -3.4604e+01 -2.5753e+01  2e+02  8e+00  7e-01
 2: -8.4823e+01 -5.7751e+01  2e+02  7e+00  6e-01
 3: -1.4835e+02 -7.4751e+01  2e+02  5e+00  5e-01
 4: -2.3702e+01 -1.2496e+01  6e+01  9e-01  8e-02
 5: -6.9865e+00 -9.3856e+00  2e+00  3e-15  2e-14
 6: -8.4536e+00 -8.4903e+00  4e-02  4e-15  1e-14
 7: -8.4811e+00 -8.4815e+00  4e-04  3e-15  1e-14
 8: -8.4814e+00 -8.4814e+00  4e-06  9e-15  1e-14
Optimal solution found.
No of. support vectors found 3
Predicted  -1   1
Truth
-1         11   0
 1          0   9
     pcost       dcost       gap    pres   dres
 0: -2.4622e+01 -4.6718e+01  5e+02  2e+01  2e+00
 1: -3.4868e+01 -2.5668e+01  2e+02  8e+00  7e-01
 2: -8.4456e+01 -5.6493e+01  2e+02  7e+00  6e-01
 3: -1.3329e+02 -6.4686e+01  2e+02  5e+00  4e-01
 4: -2.0204e+01 -1.1914e+01  5e+01  7e-01  6e-02
 5: -7.2273e+00 -9.2022e+00  2e+00  1e-14  2e-14
 6: -8.4610e+00 -8.4890e+00  3e-02  7e-15  1e-14
 7: -8.4812e+00 -8.4814e+00  3e-04  4e-15  1e-14
 8: -8.4814e+00 -8.4814e+00  3e-06  4e-15  1e-14
Optimal solution found.
No of. support vectors found 4
Predicted  -1   1
Truth
-1         10   0
 1          0  10
     pcost       dcost       gap    pres   dres
 0: -2.3390e+01 -4.4338e+01  5e+02  2e+01  2e+00
 1: -3.4870e+01 -2.5245e+01  2e+02  8e+00  7e-01
 2: -8.4451e+01 -5.6594e+01  2e+02  7e+00  6e-01
 3: -1.3777e+02 -6.9221e+01  2e+02  5e+00  5e-01
 4: -2.6201e+01 -1.2724e+01  6e+01  1e+00  9e-02
 5: -7.0348e+00 -9.4468e+00  2e+00  2e-14  2e-14
 6: -8.4535e+00 -8.4909e+00  4e-02  5e-15  1e-14
 7: -8.4811e+00 -8.4815e+00  4e-04  2e-15  1e-14
```

```
 8: -8.4814e+00 -8.4814e+00  4e-06  3e-15  9e-15
Optimal solution found.
No of. support vectors found 3
Predicted  -1   1
Truth
-1          7   0
 1          0  13
     pcost       dcost       gap    pres   dres
 0: -2.3832e+01 -4.4585e+01  5e+02  2e+01  2e+00
 1: -3.3708e+01 -2.2896e+01  2e+02  7e+00  6e-01
 2: -4.1892e+01 -2.5489e+01  2e+02  6e+00  5e-01
 3: -7.0550e+01 -2.0084e+01  6e+01  3e+00  2e-01
 4: -3.8647e+00 -6.1422e+00  4e+00  2e-02  2e-03
 5: -5.3368e+00 -5.4237e+00  1e-01  1e-03  8e-05
 6: -5.4134e+00 -5.4143e+00  1e-03  1e-05  9e-07
 7: -5.4142e+00 -5.4142e+00  1e-05  1e-07  9e-09
 8: -5.4142e+00 -5.4142e+00  1e-07  1e-09  9e-11
Optimal solution found.
No of. support vectors found 3
Predicted  -1   1
Truth
-1         10   0
 1          0  10
     pcost       dcost       gap    pres   dres
 0: -2.4651e+01 -4.6811e+01  5e+02  2e+01  2e+00
 1: -3.5232e+01 -2.4813e+01  2e+02  8e+00  6e-01
 2: -8.3499e+01 -5.4001e+01  2e+02  7e+00  6e-01
 3: -1.1418e+02 -5.7401e+01  2e+02  5e+00  4e-01
 4: -2.9385e+01 -1.2469e+01  6e+01  1e+00  9e-02
 5: -6.9884e+00 -9.1444e+00  2e+00  8e-15  2e-14
 6: -8.2068e+00 -8.2387e+00  3e-02  4e-15  1e-14
 7: -8.2294e+00 -8.2297e+00  3e-04  6e-15  1e-14
 8: -8.2296e+00 -8.2296e+00  3e-06  1e-15  1e-14
Optimal solution found.
No of. support vectors found 3
Predicted  -1   1
Truth
-1          7   0
 1          0  13
################################
Average of the model parameters
################################
Error rate: 0.005
Accuracy: 0.995
Precision: [ 1.          0.98888889]
Recall: [ 1.          0.98888889]
F-measure: [ 1.          0.98888889]
```

Plot for the data points and the support vectors:



The ***yellow circles*** represent the support vectors.

Results observed for non-linearly separable data (sample results):

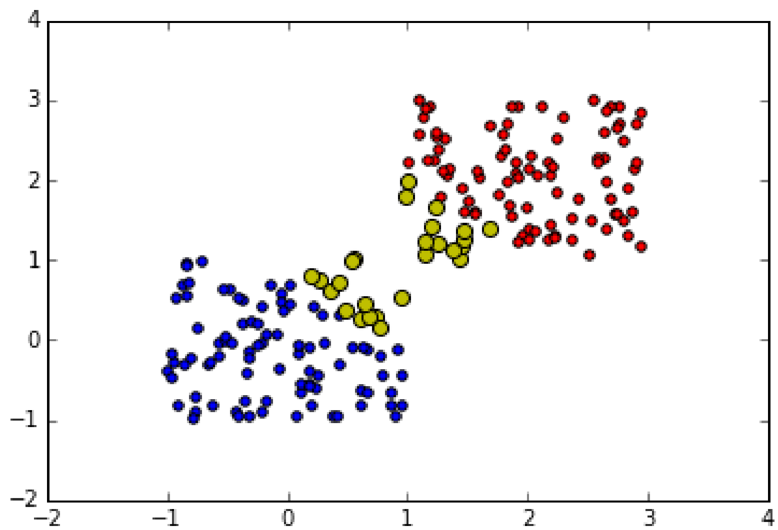
```
pcost       dcost       gap    pres   dres
 0: -8.8577e+01 -2.2716e+02  7e+02  2e+01  3e+00
 1: -3.1782e+02 -5.7778e+02  6e+02  2e+01  2e+00
 2: -1.3079e+03 -1.7929e+03  5e+02  1e+01  1e+00
 3: -3.2327e+03 -3.9612e+03  8e+02  1e+01  1e+00
 4: -4.3670e+03 -5.2889e+03  9e+02  1e+01  1e+00
 5: -9.7243e+03 -1.1335e+04  2e+03  1e+01  1e+00
 6: -1.0102e+04 -1.1766e+04  2e+03  1e+01  1e+00
 7: -3.8779e+04 -4.2575e+04  4e+03  9e+00  1e+00
 8: -2.1971e+05 -2.3222e+05  1e+04  9e+00  1e+00
 9: -2.8187e+06 -2.8767e+06  6e+04  9e+00  1e+00
10: -1.2621e+08 -1.2652e+08  3e+05  8e+00  1e+00
11: -4.9801e+10 -4.9807e+10  6e+06  8e+00  1e+00
12: -6.5378e+10 -6.5386e+10  8e+06  8e+00  1e+00
13: -6.5385e+10 -6.5393e+10  8e+06  8e+00  1e+00
14: -6.5255e+10 -6.5263e+10  8e+06  8e+00  1e+00
15: -7.3081e+10 -7.3090e+10  9e+06  8e+00  1e+00
16: -8.4575e+10 -8.4584e+10  9e+06  8e+00  1e+00
17: -9.5457e+10 -9.5466e+10  9e+06  8e+00  1e+00
18: -1.3777e+11 -1.3778e+11  1e+07  8e+00  1e+00
Terminated (singular KKT matrix).
No of. support vectors found 180
Predicted  -1   1
Truth
-1          7   2
```

```
 1          0  11
     pcost       dcost       gap    pres   dres
 0: -8.8666e+01 -2.2740e+02  7e+02  2e+01  3e+00
 1: -3.5906e+02 -6.2510e+02  5e+02  1e+01  2e+00
 2: -1.5007e+03 -1.9646e+03  5e+02  1e+01  1e+00
 3: -4.6937e+03 -5.4444e+03  8e+02  1e+01  1e+00
 4: -2.5066e+04 -2.6685e+04  2e+03  9e+00  1e+00
 5: -1.1996e+05 -1.2593e+05  6e+03  9e+00  1e+00
 6: -6.5122e+05 -6.7682e+05  3e+04  9e+00  1e+00
 7: -7.8018e+06 -7.9809e+06  2e+05  9e+00  1e+00
 8: -2.8993e+08 -2.9118e+08  1e+06  9e+00  1e+00
 9: -6.6025e+10 -6.6032e+10  7e+06  9e+00  1e+00
10: -1.1363e+11 -1.1364e+11  1e+07  9e+00  1e+00
Terminated (singular KKT matrix).
No of. support vectors found 180
Predicted  -1   1
Truth
-1         10   3
 1          0   7
     pcost       dcost       gap    pres   dres
 0: -8.6349e+01 -2.2405e+02  7e+02  2e+01  3e+00
 1: -3.1092e+02 -5.7036e+02  6e+02  2e+01  2e+00
 2: -1.3163e+03 -1.8042e+03  5e+02  1e+01  1e+00
 3: -3.1689e+03 -3.9002e+03  8e+02  1e+01  1e+00
 4: -4.4945e+03 -5.4520e+03  1e+03  1e+01  1e+00
 5: -1.1075e+04 -1.2822e+04  2e+03  1e+01  1e+00
 6: -1.2410e+04 -1.4335e+04  2e+03  1e+01  1e+00
 7: -6.5650e+04 -6.9903e+04  4e+03  9e+00  1e+00
 8: -7.3794e+05 -7.5589e+05  2e+04  8e+00  1e+00
 9: -2.7584e+07 -2.7675e+07  9e+04  8e+00  1e+00
10: -8.0791e+09 -8.0802e+09  1e+06  8e+00  1e+00
11: -5.9532e+10 -5.9540e+10  8e+06  8e+00  1e+00
12: -5.9961e+10 -5.9969e+10  8e+06  8e+00  1e+00
13: -7.6318e+10 -7.6328e+10  1e+07  8e+00  1e+00
Terminated (singular KKT matrix).
No of. support vectors found 180
Predicted  -1   1
Truth
-1         10   0
 1          3   7
################################
Average of the model parameters
################################
Error rate: 0.16
Accuracy: 0.84
Precision: [ 0.85199523  0.83740065]
Recall: [ 0.85199523  0.83740065]
F-measure: [ 0.85199523  0.83740065]
```

From the above results, it is observed that there is a decline with the performance of the algorithm and optimal solution could not be found for certain cases. This shows that linear SVM doesn't do well for non-linear data. Hence we should go for SVMs that do well for non-linear data.

4) Linear SVM with soft margin:

```
Predicted  -1   1
Truth
-1          9   0
 1          0  11
Predicted  -1   1
Truth
-1         13   0
 1          0   7
Predicted  -1   1
Truth
-1         10   0
 1          0  10
Predicted  -1   1
Truth
-1         11   0
 1          0   9
Predicted  -1   1
Truth
-1         11   1
 1          0   8
################################
Average of the model parameters
################################
Error rate: 0.01
Accuracy: 0.99
Precision: [ 1.          0.98174603]
Recall: [ 1.          0.98174603]
F-measure: [ 1.          0.98174603]
```



The yellow circles represent the support vectors.

Non-linear data:

```
################################
Average of the model parameters
################################
Error rate: 0.165
Accuracy: 0.835
Precision: [ 0.88445554  0.80781441]
Recall: [ 0.88445554  0.80781441]
F-measure: [ 0.88445554  0.80781441]
```

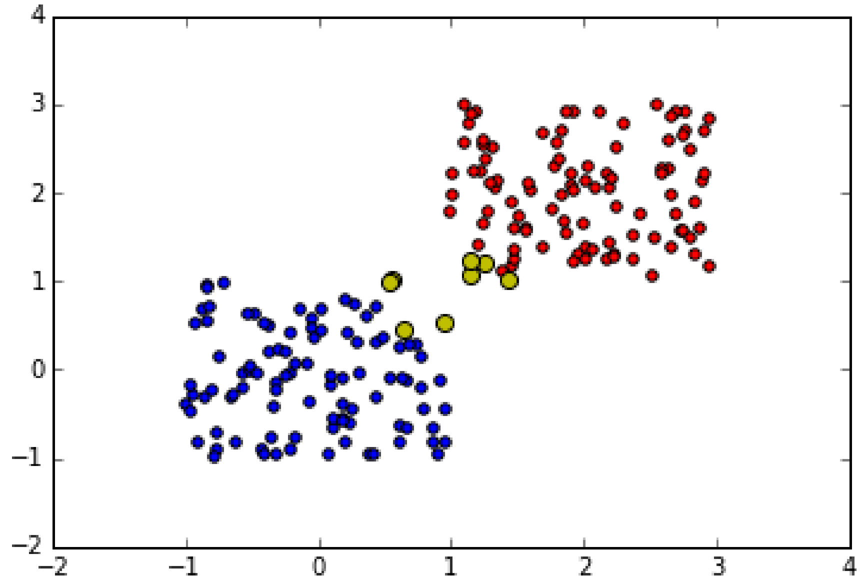
5) Kernel-based SVM:

Polynomial and Gaussian kernels are implemented for the SVM and below are the results. From the above plot for the non-linearly separable data, the data points intersect with each other and it is difficult to define a decision boundary that perfectly segregates the classes. So we depend on polynomial and Gaussian kernels to project the data into higher dimensional space.

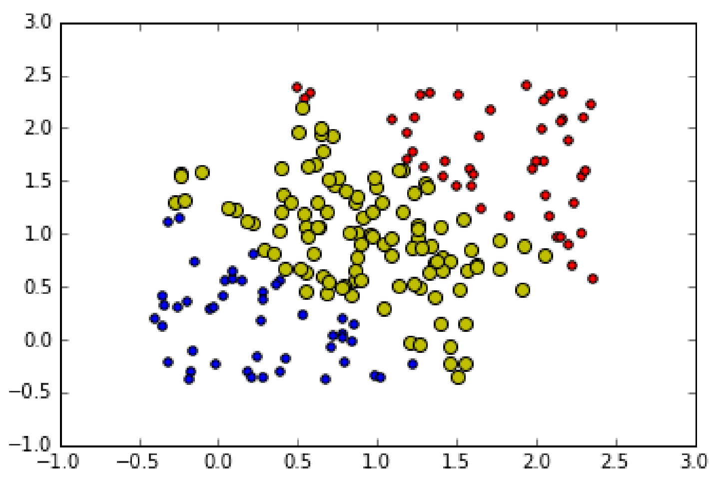
Kernels		Accuracy	
Polynomial(degree)	Gaussian(sigma)	Polynomial	Gaussian
2	1.0	0.85	0.83
3	2.0	0.82	0.83
4	0.2	0.73	0.835
5	0.5	0.61	0.84

From the above results, polynomial of degree 2 and Gaussian kernel with sigma 0.5 does well in classifying the examples. Also, by changing the slack values an improved accuracy of 85% was obtained with a slack 0f 0.2 and sigma >10.

Support vectors for linear data:

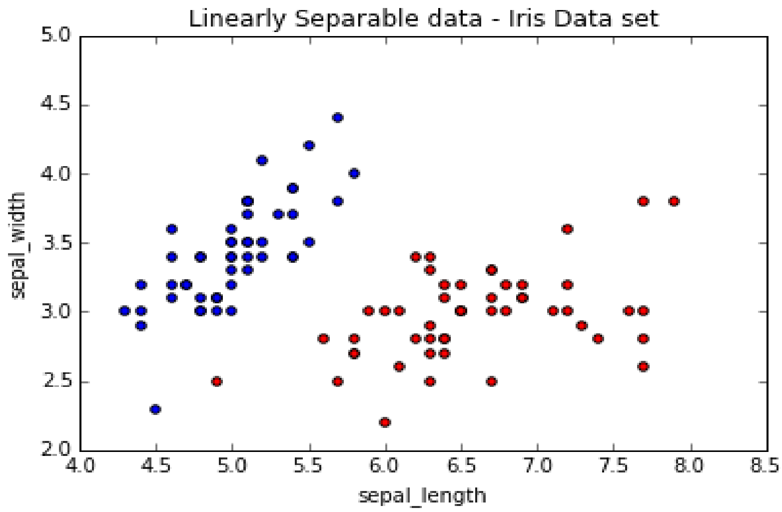


Example of how it looks to be support vectors for non-linearly separable data as it is not easy to identify the support vetors.



External Datasets:

The algorithm is verified for its correctness on external data sets such as Iris and Diabates. Following are the results.



Above is the plot for iris dataset for the features sepal width and sepal length and, setosa and virginica as the classes.

Results obtained are given below:

SVM	Parameters	Accuracy
Polynomial	2,3	0.97,0.93
Gaussian	0.5,0.4	0.87,0.83

Similar to the above, the experiment worked well with a polynomial of degree 2 and sigma of 0.5.

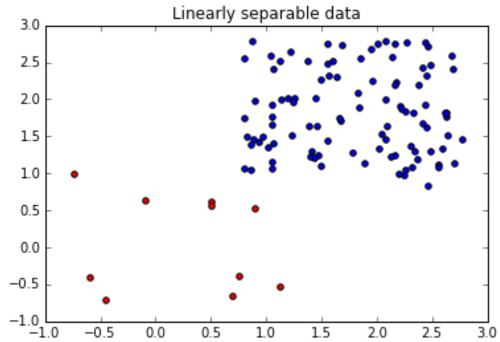
6) Let us consider the case when class ‘1’ of the generated dataset has more examples than class ‘-1’. For eg., class 1 has 100 samples and class ‘-1’ has 10 samples.

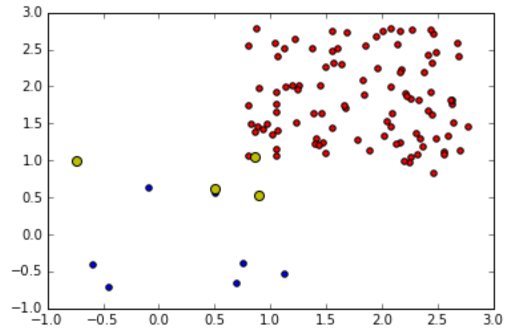
By running the experiment using hard margin the below results are observed:

```
################################
Average of the model parameters
################################
Error rate: 0.372727272727
Accuracy: 0.627272727273
Precision: [ 0.49333333  1.115     ]
Recall: [ 0.49333333  1.115     ]
F-measure: [ 0.49333333  1.115     ]
```

This is because with the current setting identification of the support vectors becomes tedious and the misclassification rate tends to be higher. A simple solution to this would be bring the data points close together so that they are still linearly separable and support vectors are identifiable

Let's take a look at the plots below:





From the above plot, we can conclude that the model does well even though there is an imbalance in the number of examples by tweaking the model parameters and maximum efficiency can be obtained.

3)

Given:-

$$\mathcal{L}_p = \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{m} \varepsilon_{e_i} - \sum_{i=1}^{m} \alpha_i \left(y^{(i)}(w^T x_i + w_0) - 1 + \varepsilon_{e_i}\right) - \sum_{i=1}^{m} \beta_i \varepsilon_{e_i} \rightarrow (1)$$

Minimization equations:-

$$\frac{\partial L_p}{\partial w} = 0, \quad \frac{\partial L_p}{\partial w_0} = 0, \quad \frac{\partial L_p}{\partial \varepsilon_{e_i}} = 0$$

$$\frac{\partial L_p}{\partial w} = 2 \times \frac{1}{2} w + 0 - \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} = 0$$

$$\Rightarrow \boxed{w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}} \rightarrow (2)$$

$$\frac{\partial L_p}{\partial w_0} = 0 + 0 - \sum_{i=1}^{m} \alpha_i y^{(i)} = 0$$

$$\Rightarrow \boxed{\sum_{i=1}^{m} \alpha_i y^{(i)} = 0} \rightarrow (3)$$

$$\frac{\partial L_p}{\partial \varepsilon_{e_i}} = c - \alpha_i - \beta_i = 0 \rightarrow (4)$$

Substituting (2), (3), (4) in (1), we get,

$$\mathcal{L}_p = \frac{1}{2}\left(\sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)T}\right)\left(\sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}\right) + c\sum_{i=1}^{m} \varepsilon_{e_i} - \sum_{i=1}^{m} \alpha_i y^{(i)} \left(\left(\sum_{i=1}^{m} \alpha_i y^{(i)} \cdot x^{(i)T}\right) x^{(i)}\right) + \sum_{i=1}^{m} \alpha_i y^{(i)} w_0 + \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \alpha_i \varepsilon_{e_i} + \sum_{i=1}^{m} \beta_i \varepsilon_{e_i} \rightarrow (5)$$

From ③, ④ we know that,

$$\sum_{i=1}^{m} \alpha_i y^{(i)} = 0 \quad \text{and} \quad C - \beta_i - \varepsilon_i = 0.$$

$$\sum_{i=1}^{m} \varepsilon_{i} (C - \beta_i - \alpha_i) = 0.$$

So,

$$\boxed{L_D = -\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} \cdot x^{(i)T} x^{(j)} + \sum_{i=1}^{m} \alpha_i}$$