# Final Project

**Weilan Yang**

May 4, 2016

## 1

The original problem

$$\begin{aligned} & min_{w,\gamma,y,z} \quad (\frac{1}{m}e^T y + \frac{1}{k}e^T z) + \frac{\mu}{2}w^T w \\ & s.t. \quad Mw - e\gamma + y \geq e, -Bw + e\gamma + z \geq e, y \geq 0, z \geq 0 \end{aligned}$$

can be formulated as a quadratic programming

$$\begin{aligned} & min \quad \frac{1}{2}x^T Q x + p^T x \\ & s.t. \quad Ax \leq b;\ y,z \geq 0;\ w,\gamma \text{ free} \end{aligned}$$

where

$$x = \begin{bmatrix} w_{n\times 1} \\ \gamma \\ y_{m\times 1} \\ z_{k\times 1} \end{bmatrix}, \quad Q = \begin{bmatrix} \mu & & & & \\ & \ddots & & & \\ & & \mu & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}, \quad p = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1/m \\ \vdots \\ 1/m \\ 1/k \\ \vdots \\ 1/k \end{bmatrix}, \quad A = \begin{bmatrix} -M & e_{m\times 1} & -I_m & 0_{m\times k} \\ B & -e_{k\times 1} & 0_{k\times m} & -I_k \end{bmatrix}, \quad b = \begin{bmatrix} -e_{(m+k)\times 1} \end{bmatrix}$$

Let's first define a function `fitModel` as below

```
function [w gamma obj] = fitModel(mu, Mtrain, Btrain)
% syntax: [w gamma] = fitModel(mu, Mtrain, Btrain)
% given the tuning param, training data, fit the model and output coeffs and
% the objective value for the QP.

% formulate the QP
n = size(Btrain,2);
m = size(Mtrain,1);
k = size(Btrain,1);
Q = [mu*eye(n) zeros(n,1+m+k); zeros(1+m+k,n+1+m+k)];
p = [zeros(n+1,1); (1/m)*ones(m,1); (1/k)*ones(k,1)];
A = [-Mtrain ones(m,1) -eye(m) zeros(m,k);
     Btrain -ones(k,1) zeros(k,m) -eye(k)];
b = -ones(m+k,1);
lb = [-inf((n+1),1); zeros(m+k,1)];
ub = inf(n+1+m+k,1);
```

```
% solve the QP
[x obj] = cplexqp(Q,p,A,b,[],[],lb,ub);
w = x(1:n);
gamma = x(n+1);
end
```

Let the tuning parameter $\mu = 0.0001$. The following is the code and output shows the solved $w$ and $\gamma$.

```
[train, tune, test] = getdata('wdbc.data', 30);
Btrain = train(find(train(:,1) == 66), 2:31);
Mtrain = train(find(train(:,1) == 77), 2:31);
mu = 0.0001;
% fit the model
[w gamma obj] = fitModel(mu, Mtrain, Btrain)

w =

   -4.1056
   -0.1242
   -3.3016
   -1.6735
    2.4501
   -4.8508
   -0.3210
    3.7928
   -1.2820
    0.2479
    4.1126
   -0.8100
   -0.3688
    3.8027
    0.5084
   -0.0471
   -0.1307
    4.2478
    0.7133
   -7.1841
    6.5976
    5.0381
    4.1452
    8.0556
   -1.6099
    0.1599
   -0.3038
    0.6853
    0.6642
    6.4163


gamma =

   -3.5688
```

## 2.

To pick the best trade off parameter, we should count the number of missclassified cases in the tuning set. Meanwhile, we should break the tie by considering the error of missclassification. The error for the missclassified cases in the tuning set could be defined as

$$\frac{1}{M}\sum_{i\in\mathcal{M}}|wx^i-\gamma|I(wx^i-\gamma\leq 0)+\frac{1}{B}\sum_{i\in\mathcal{B}}|wx^i-\gamma|I(wx^i-\gamma>0)$$

where $I(\cdot)$ is a indicator function.

Now we can define a function `evaluate` as below, for picking up the best $\mu$.

```
function [numMissed errors] = evaluate(mues, Mtrain, Btrain, Mtest, Btest)
% syntax: missed = getMissed(mu, Mtrain, Btrain, Mtest, Btest)
% given a vector of tuning parameter, training sets for positive and
% negative examples, test sets for positive and negative examples
% (features only, no class labels), fit the svm model and evaluate
% the model accuracy on test data.

numMissed = [];
errors = [];
for idx = 1:numel(mues)
  mu = mues(idx);
  [w gamma obj] = fitModel(mu, Mtrain, Btrain);
  % evaluate accuracy on the test data
  predictM = Mtest * w - gamma;
  predictB = Btest * w - gamma;
  correctM = (predictM > 0);
  wrongM = (predictM <= 0);
  correctB = (predictB <= 0);
  wrongB = (predictB > 0);
  numMissed = [numMissed sum(wrongM)+sum(wrongB)];
  m = size(predictM,1);
  k = size(predictB,1);
  err = (1/m)*sum(abs(predictM(find(wrongM == 1))));
  err = err + (1/k)*sum(abs(predictB(find(wrongB == 1))));
  errors = [errors err];
end
end
```

Using the function, we can investigate the number of missclassified cases in tuning set for each $\mu$'s.

```
Btune = tune(find(tune(:,1) == 66), 2:31);
Mtune = tune(find(tune(:,1) == 77), 2:31);
% evaluate the model when mu = 0.0001
[missed obj] = evaluate(mu, Mtrain, Btrain, Mtune, Btune)

missed =

     3


obj =
```

```
    0.1670

% evaluate the model for different mu's
mues = 5e-5:5e-5:5e-4;
[missed errors] = evaluate(mues, Mtrain, Btrain, Mtune, Btune)

missed =

     3     3     2     2     2     2     2     2     2     2


errors =

  Columns 1 through 8

    0.2763    0.1670    0.1313    0.1054    0.0925    0.0867    0.0820    0.0673

  Columns 9 through 10

    0.0534    0.0468
```

The number of missclassified points are all 2 for $\mu = 1.5e-4, \cdots, 5e-4$. We can break the tie by the error on the tuning set. When $\mu = 5e-4$ the error is the smallest, which is 0.0468. So, the best $\mu = 5e-4$.

# 3.

```
mu = 5e-4;
minMissed = size(Mtune,1);
for i = 1:29
  for j = (i+1):30
    Msubtrain = Mtrain(:,[i j]);
    Bsubtrain = Btrain(:,[i j]);
    Msubtune = Mtune(:,[i j]);
    Bsubtune = Btune(:,[i j]);
    numMissed = evaluate(mu,Msubtrain,Bsubtrain,Msubtune,Bsubtune);
    if numMissed < minMissed
      minMissed = numMissed;
    end
    fprintf('atts %2d %2d: misclass %3d\n',i,j, numMissed);
  end
end
minMissed

atts  1  2: misclass  14
atts  1  3: misclass  11
atts  1  4: misclass  12
atts  1  5: misclass  10
atts  1  6: misclass   8
atts  1  7: misclass   8
atts  1  8: misclass   8
atts  1  9: misclass   9
atts  1 10: misclass  11
atts  1 11: misclass   9
atts  1 12: misclass  14
atts  1 13: misclass   8
atts  1 14: misclass   8
atts  1 15: misclass  12
atts  1 16: misclass  11
atts  1 17: misclass  12
atts  1 18: misclass  11
atts  1 19: misclass  12
atts  1 20: misclass  12
atts  1 21: misclass   5
atts  1 22: misclass  13
atts  1 23: misclass   8
atts  1 24: misclass   6
atts  1 25: misclass   5
atts  1 26: misclass  12
atts  1 27: misclass  11
atts  1 28: misclass   8
atts  1 29: misclass   9
atts  1 30: misclass   9
atts  2  3: misclass  13
atts  2  4: misclass  13
atts  2  5: misclass  22
atts  2  6: misclass  25
```

```
atts  2  7: misclass  18
atts  2  8: misclass   7
atts  2  9: misclass  28
atts  2 10: misclass  32
atts  2 11: misclass  20
atts  2 12: misclass  30
atts  2 13: misclass  23
atts  2 14: misclass  14
atts  2 15: misclass  31
atts  2 16: misclass  26
atts  2 17: misclass  28
atts  2 18: misclass  27
atts  2 19: misclass  32
atts  2 20: misclass  32
atts  2 21: misclass  11
atts  2 22: misclass  32
atts  2 23: misclass  10
atts  2 24: misclass  11
atts  2 25: misclass  18
atts  2 26: misclass  26
atts  2 27: misclass  19
atts  2 28: misclass  12
atts  2 29: misclass  27
atts  2 30: misclass  27
atts  3  4: misclass  10
atts  3  5: misclass   9
atts  3  6: misclass   9
atts  3  7: misclass  10
atts  3  8: misclass   9
atts  3  9: misclass   9
atts  3 10: misclass  10
atts  3 11: misclass  11
atts  3 12: misclass  13
atts  3 13: misclass  10
atts  3 14: misclass   6
atts  3 15: misclass  10
atts  3 16: misclass  12
atts  3 17: misclass  11
atts  3 18: misclass  11
atts  3 19: misclass  11
atts  3 20: misclass  11
atts  3 21: misclass   9
atts  3 22: misclass  12
atts  3 23: misclass   7
atts  3 24: misclass   7
atts  3 25: misclass   5
atts  3 26: misclass  11
atts  3 27: misclass  11
atts  3 28: misclass   8
atts  3 29: misclass   9
atts  3 30: misclass   9
atts  4  5: misclass   9
```

```
atts  4  6: misclass   7
atts  4  7: misclass   7
atts  4  8: misclass   8
atts  4  9: misclass   9
atts  4 10: misclass  10
atts  4 11: misclass   8
atts  4 12: misclass  14
atts  4 13: misclass   7
atts  4 14: misclass   7
atts  4 15: misclass  13
atts  4 16: misclass  11
atts  4 17: misclass  12
atts  4 18: misclass  11
atts  4 19: misclass  12
atts  4 20: misclass  12
atts  4 21: misclass   8
atts  4 22: misclass  13
atts  4 23: misclass   8
atts  4 24: misclass   6
atts  4 25: misclass   5
atts  4 26: misclass  11
atts  4 27: misclass  11
atts  4 28: misclass   8
atts  4 29: misclass   9
atts  4 30: misclass   9
atts  5  6: misclass  18
atts  5  7: misclass  12
atts  5  8: misclass   8
atts  5  9: misclass  29
atts  5 10: misclass  23
atts  5 11: misclass  13
atts  5 12: misclass  25
atts  5 13: misclass  13
atts  5 14: misclass   7
atts  5 15: misclass  24
atts  5 16: misclass  28
atts  5 17: misclass  22
atts  5 18: misclass  18
atts  5 19: misclass  27
atts  5 20: misclass  27
atts  5 21: misclass   6
atts  5 22: misclass  23
atts  5 23: misclass   8
atts  5 24: misclass   6
atts  5 25: misclass  28
atts  5 26: misclass  24
atts  5 27: misclass  18
atts  5 28: misclass   9
atts  5 29: misclass  27
atts  5 30: misclass  26
atts  6  7: misclass  10
atts  6  8: misclass   8
```

```
atts  6  9: misclass  17
atts  6 10: misclass  14
atts  6 11: misclass   8
atts  6 12: misclass  17
atts  6 13: misclass  12
atts  6 14: misclass   5
atts  6 15: misclass  19
atts  6 16: misclass  15
atts  6 17: misclass  13
atts  6 18: misclass  19
atts  6 19: misclass  16
atts  6 20: misclass  15
atts  6 21: misclass   8
atts  6 22: misclass  24
atts  6 23: misclass   7
atts  6 24: misclass   8
atts  6 25: misclass  14
atts  6 26: misclass  19
atts  6 27: misclass  19
atts  6 28: misclass   9
atts  6 29: misclass  17
atts  6 30: misclass  14
atts  7  8: misclass   8
atts  7  9: misclass  12
atts  7 10: misclass   9
atts  7 11: misclass   6
atts  7 12: misclass  12
atts  7 13: misclass   7
atts  7 14: misclass   5
atts  7 15: misclass  12
atts  7 16: misclass   9
atts  7 17: misclass   9
atts  7 18: misclass   8
atts  7 19: misclass  10
atts  7 20: misclass   9
atts  7 21: misclass   7
atts  7 22: misclass  19
atts  7 23: misclass   8
atts  7 24: misclass   7
atts  7 25: misclass  11
atts  7 26: misclass  13
atts  7 27: misclass  12
atts  7 28: misclass  10
atts  7 29: misclass   9
atts  7 30: misclass  10
atts  8  9: misclass   8
atts  8 10: misclass   8
atts  8 11: misclass   7
atts  8 12: misclass   8
atts  8 13: misclass   7
atts  8 14: misclass   3
atts  8 15: misclass  10
```

```
atts  8 16: misclass   7
atts  8 17: misclass   8
atts  8 18: misclass   8
atts  8 19: misclass   9
atts  8 20: misclass   9
atts  8 21: misclass   6
atts  8 22: misclass   8
atts  8 23: misclass   8
atts  8 24: misclass   6
atts  8 25: misclass   7
atts  8 26: misclass   8
atts  8 27: misclass   8
atts  8 28: misclass   8
atts  8 29: misclass   7
atts  8 30: misclass   8
atts  9 10: misclass  35
atts  9 11: misclass  12
atts  9 12: misclass  39
atts  9 13: misclass  14
atts  9 14: misclass   7
atts  9 15: misclass  39
atts  9 16: misclass  35
atts  9 17: misclass  35
atts  9 18: misclass  27
atts  9 19: misclass  41
atts  9 20: misclass  35
atts  9 21: misclass   7
atts  9 22: misclass  28
atts  9 23: misclass   8
atts  9 24: misclass   8
atts  9 25: misclass  24
atts  9 26: misclass  21
atts  9 27: misclass  18
atts  9 28: misclass   9
atts  9 29: misclass  33
atts  9 30: misclass  28
atts 10 11: misclass  12
atts 10 12: misclass  75
atts 10 13: misclass   9
atts 10 14: misclass   7
atts 10 15: misclass  69
atts 10 16: misclass  27
atts 10 17: misclass  28
atts 10 18: misclass  23
atts 10 19: misclass  70
atts 10 20: misclass  31
atts 10 21: misclass   7
atts 10 22: misclass  33
atts 10 23: misclass   8
atts 10 24: misclass   6
atts 10 25: misclass  25
atts 10 26: misclass  18
```

```
atts 10 27: misclass  16
atts 10 28: misclass   7
atts 10 29: misclass  33
atts 10 30: misclass  23
atts 11 12: misclass  14
atts 11 13: misclass  12
atts 11 14: misclass   5
atts 11 15: misclass  12
atts 11 16: misclass  12
atts 11 17: misclass  12
atts 11 18: misclass  12
atts 11 19: misclass  12
atts 11 20: misclass  13
atts 11 21: misclass   7
atts 11 22: misclass  21
atts 11 23: misclass   8
atts 11 24: misclass   8
atts 11 25: misclass   9
atts 11 26: misclass  16
atts 11 27: misclass   7
atts 11 28: misclass   6
atts 11 29: misclass  13
atts 11 30: misclass  15
atts 12 13: misclass  14
atts 12 14: misclass   9
atts 12 15: misclass  69
atts 12 16: misclass  30
atts 12 17: misclass  30
atts 12 18: misclass  25
atts 12 19: misclass  69
atts 12 20: misclass  34
atts 12 21: misclass   7
atts 12 22: misclass  32
atts 12 23: misclass   7
atts 12 24: misclass   8
atts 12 25: misclass  24
atts 12 26: misclass  21
atts 12 27: misclass  19
atts 12 28: misclass   8
atts 12 29: misclass  39
atts 12 30: misclass  33
atts 13 14: misclass   6
atts 13 15: misclass  10
atts 13 16: misclass  10
atts 13 17: misclass  10
atts 13 18: misclass  10
atts 13 19: misclass  15
atts 13 20: misclass  12
atts 13 21: misclass   6
atts 13 22: misclass  23
atts 13 23: misclass   8
atts 13 24: misclass   8
```

```
atts 13 25: misclass  12
atts 13 26: misclass  16
atts 13 27: misclass  12
atts 13 28: misclass   8
atts 13 29: misclass  12
atts 13 30: misclass  15
atts 14 15: misclass   7
atts 14 16: misclass   7
atts 14 17: misclass   7
atts 14 18: misclass   7
atts 14 19: misclass   6
atts 14 20: misclass   6
atts 14 21: misclass   9
atts 14 22: misclass  21
atts 14 23: misclass   8
atts 14 24: misclass   8
atts 14 25: misclass   9
atts 14 26: misclass   8
atts 14 27: misclass   8
atts 14 28: misclass   5
atts 14 29: misclass   8
atts 14 30: misclass   8
atts 15 16: misclass  31
atts 15 17: misclass  37
atts 15 18: misclass  29
atts 15 19: misclass  67
atts 15 20: misclass  38
atts 15 21: misclass   8
atts 15 22: misclass  32
atts 15 23: misclass   6
atts 15 24: misclass   8
atts 15 25: misclass  18
atts 15 26: misclass  22
atts 15 27: misclass  18
atts 15 28: misclass   9
atts 15 29: misclass  31
atts 15 30: misclass  30
atts 16 17: misclass  28
atts 16 18: misclass  22
atts 16 19: misclass  31
atts 16 20: misclass  33
atts 16 21: misclass   8
atts 16 22: misclass  35
atts 16 23: misclass   8
atts 16 24: misclass   8
atts 16 25: misclass  25
atts 16 26: misclass  20
atts 16 27: misclass  15
atts 16 28: misclass   8
atts 16 29: misclass  29
atts 16 30: misclass  25
atts 17 18: misclass  24
```

```
atts 17 19: misclass  27
atts 17 20: misclass  27
atts 17 21: misclass  11
atts 17 22: misclass  34
atts 17 23: misclass   8
atts 17 24: misclass  10
atts 17 25: misclass  22
atts 17 26: misclass  20
atts 17 27: misclass  17
atts 17 28: misclass   9
atts 17 29: misclass  28
atts 17 30: misclass  26
atts 18 19: misclass  19
atts 18 20: misclass  19
atts 18 21: misclass   6
atts 18 22: misclass  28
atts 18 23: misclass   7
atts 18 24: misclass   7
atts 18 25: misclass  15
atts 18 26: misclass  20
atts 18 27: misclass  18
atts 18 28: misclass   9
atts 18 29: misclass  22
atts 18 30: misclass  23
atts 19 20: misclass  33
atts 19 21: misclass   8
atts 19 22: misclass  33
atts 19 23: misclass   9
atts 19 24: misclass   8
atts 19 25: misclass  27
atts 19 26: misclass  22
atts 19 27: misclass  18
atts 19 28: misclass   9
atts 19 29: misclass  33
atts 19 30: misclass  32
atts 20 21: misclass  10
atts 20 22: misclass  33
atts 20 23: misclass   8
atts 20 24: misclass  10
atts 20 25: misclass  28
atts 20 26: misclass  19
atts 20 27: misclass  18
atts 20 28: misclass   7
atts 20 29: misclass  34
atts 20 30: misclass  29
atts 21 22: misclass   9
atts 21 23: misclass   9
atts 21 24: misclass  10
atts 21 25: misclass   5
atts 21 26: misclass   8
atts 21 27: misclass   7
atts 21 28: misclass   7
```

```
atts 21 29: misclass   7
atts 21 30: misclass   7
atts 22 23: misclass   9
atts 22 24: misclass  10
atts 22 25: misclass  24
atts 22 26: misclass  23
atts 22 27: misclass  18
atts 22 28: misclass  11
atts 22 29: misclass  28
atts 22 30: misclass  36
atts 23 24: misclass   9
atts 23 25: misclass   6
atts 23 26: misclass   9
atts 23 27: misclass   9
atts 23 28: misclass   7
atts 23 29: misclass   8
atts 23 30: misclass   7
atts 24 25: misclass   3
atts 24 26: misclass   8
atts 24 27: misclass   8
atts 24 28: misclass   6
atts 24 29: misclass   6
atts 24 30: misclass   7
atts 25 26: misclass  20
atts 25 27: misclass  18
atts 25 28: misclass   9
atts 25 29: misclass  26
atts 25 30: misclass  26
atts 26 27: misclass  19
atts 26 28: misclass   9
atts 26 29: misclass  19
atts 26 30: misclass  19
atts 27 28: misclass   9
atts 27 29: misclass  20
atts 27 30: misclass  15
atts 28 29: misclass   9
atts 28 30: misclass   8
atts 29 30: misclass  30

minMissed =

     3
```

# 4.

From the last part, we saw that the minimum number of missclassified cases in tuning set is 3. We can use the same definition of error as in part 2 to break the tie.

```
mu = 5e-4;
for i = 1:29
  for j = (i+1):30
    Msubtrain = Mtrain(:,[i j]);
    Bsubtrain = Btrain(:,[i j]);
    Msubtune = Mtune(:,[i j]);
    Bsubtune = Btune(:,[i j]);
    [numMissed err] = evaluate(mu,Msubtrain,Bsubtrain,Msubtune,Bsubtune);
    if numMissed == 3
      fprintf('atts %2d %2d: misclass %3d\t error %f\n',i,j,numMissed,err);
    end
  end
end


atts  8 14: misclass   3     error 0.052644
atts 24 25: misclass   3     error 0.028699
```

So, we can see that attribute pair (24,25) gives the least missclassified cases and the best tuning set error.

Using feature 24 and 25, we can fit the model and test on the testing set. We can see that there are 6 missclassified points.

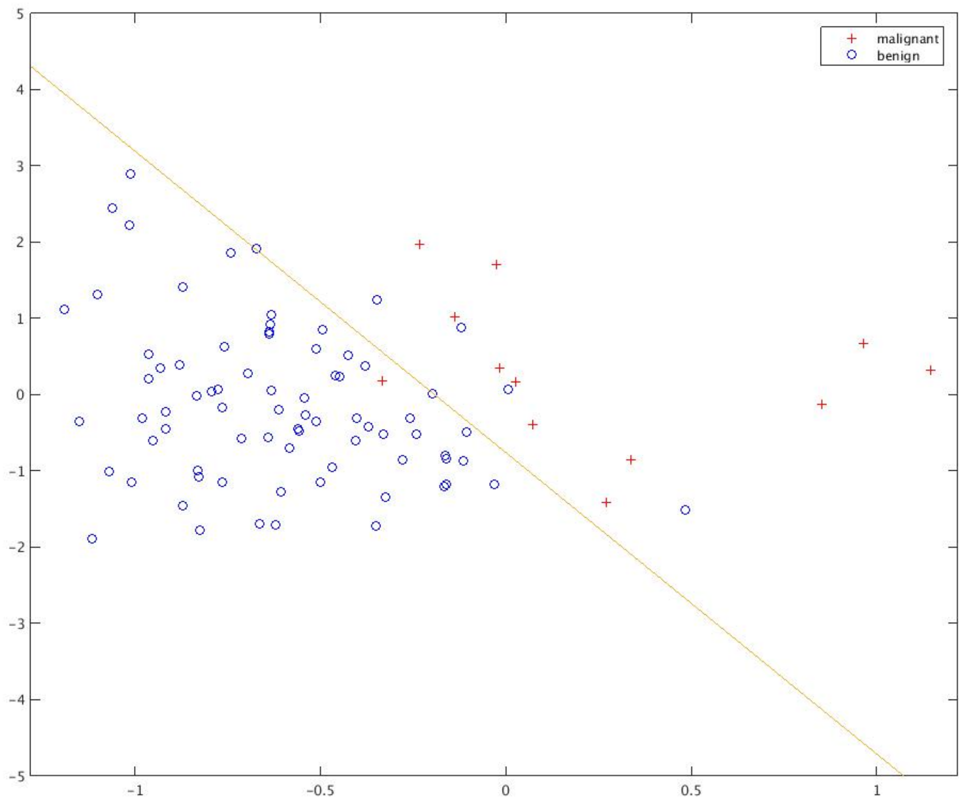
```
Btest = test(find(test(:,1) == 66), 2:31);
Mtest = test(find(test(:,1) == 77), 2:31);
mu = 5e-4;
idxs = [24 25];
Mtrain2 = Mtrain(:,idxs);
Btrain2 = Btrain(:,idxs);
Mtest2 = Mtest(:,idxs);
Btest2 = Btest(:,idxs);
numMissed = evaluate(mu,Mtrain2,Btrain2,Mtest2,Btest2)

numMissed =

     6
```

Let's now plot the testing data points, along with the separation plane.

```
% plot
[w,gamma] = fitModel(mu,Mtrain2,Btrain2);
w1 = w(1);
w2 = w(2);
xs = linspace(-2,4,1000);
ys = (-w1/w2)*xs + gamma/w2;
```

```
plot(Mtest2(:,1),Mtest2(:,2),'+r'), hold on;
plot(Btest2(:,1),Btest2(:,2),'ob'), hold on;
plot(xs,ys)
ylim([-5 5])
legend('malignant','benign')
```

The outcoming plot is shown below. It can be seen that, after some magnifying, six data points are incorrectly classified: 5 benigns are missclassified as malignant and one malignant is missclssified as benign.




separation line on feature 24 and 25, with testing data points