Homework #2

1. **Probably Approximately Correct**
   1. We know the formula , use this we can derive as a function of , steps:

Take logarithm on both sides:

Transpose to one side as a function:

(ans.)

* 1. , make a substitution to the function .

For the request,

.

Which means,

. We need 1060 examples. (ans.)

* 1. . Reuse the inequality of N.

.

We need 1981 examples. (ans.)

* 1. . Reuse inequality as above,

.

We need 2902 examples. (ans.)

1. **Gradient and Newton Directions**
   1. The first order Taylor’s expansion of E around :

.

Thus, by compare with the coefficient, we can get

; ; . (ans.)

* 1. Denote the partial derivate as , and rewrite the equation as an inequality:

Here we can prove that if we the minimum is occurred at , which mean the is parallel to , which called the negative gradient direction. (ans.)

* 1. The series to second order Taylor’s expansion of E around :

Where, .

Thus, . Compare with the coefficient, we can get

(ans.)

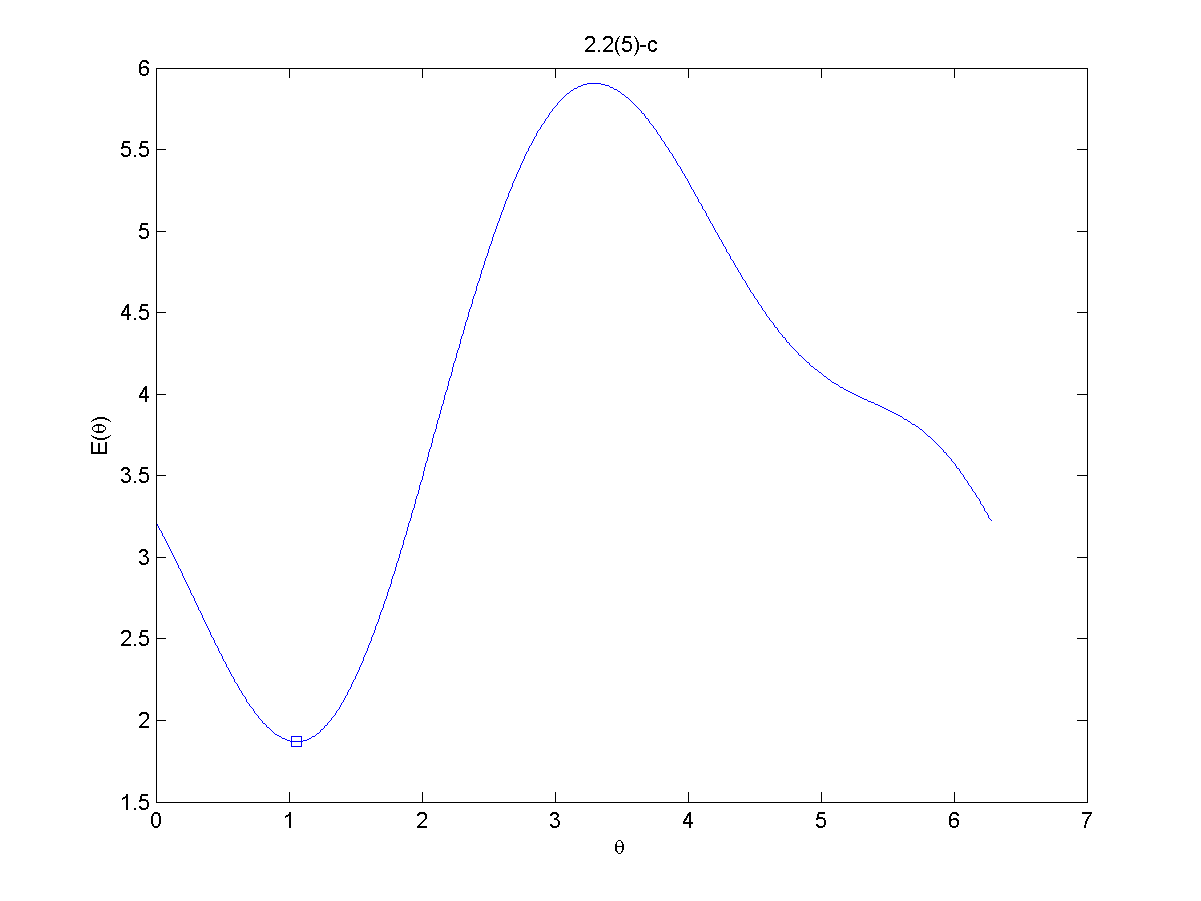
* 1. Rewrite the function as a matrix from:

To find the minimum value, we do the partial derivative on and let the result be zero. Because is positive define, we have a symmetry matrix .

Same reason as above, the determine is positive and imply we have an inverse matrix of , we get the Newton direction:

* 1. Computing the results:
     1. , function value
     2. , function value
     3. We use MATLAB to approach the answer, consider we lied on a circle with diameter equal to 1. Thus, we can change the function to domain, as below:

We compute all possibility lied on , get the figure below:



The minimum has showed with the square symbol. The position of the minimum is about . The minimum is smaller than two approach ( (a) and (b) )above.

1. **Least-Square Linear Regression**

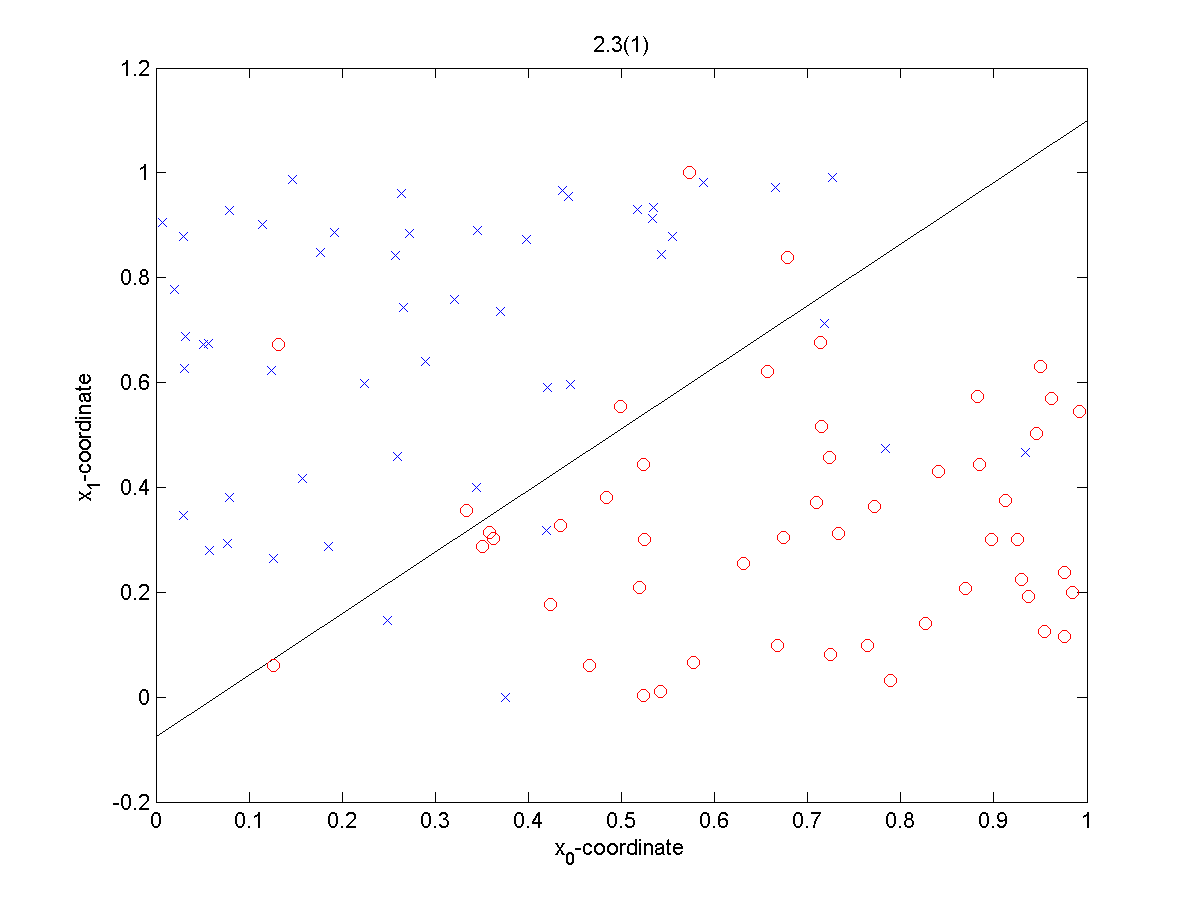


Figure 1. A least-square regression success separate the data

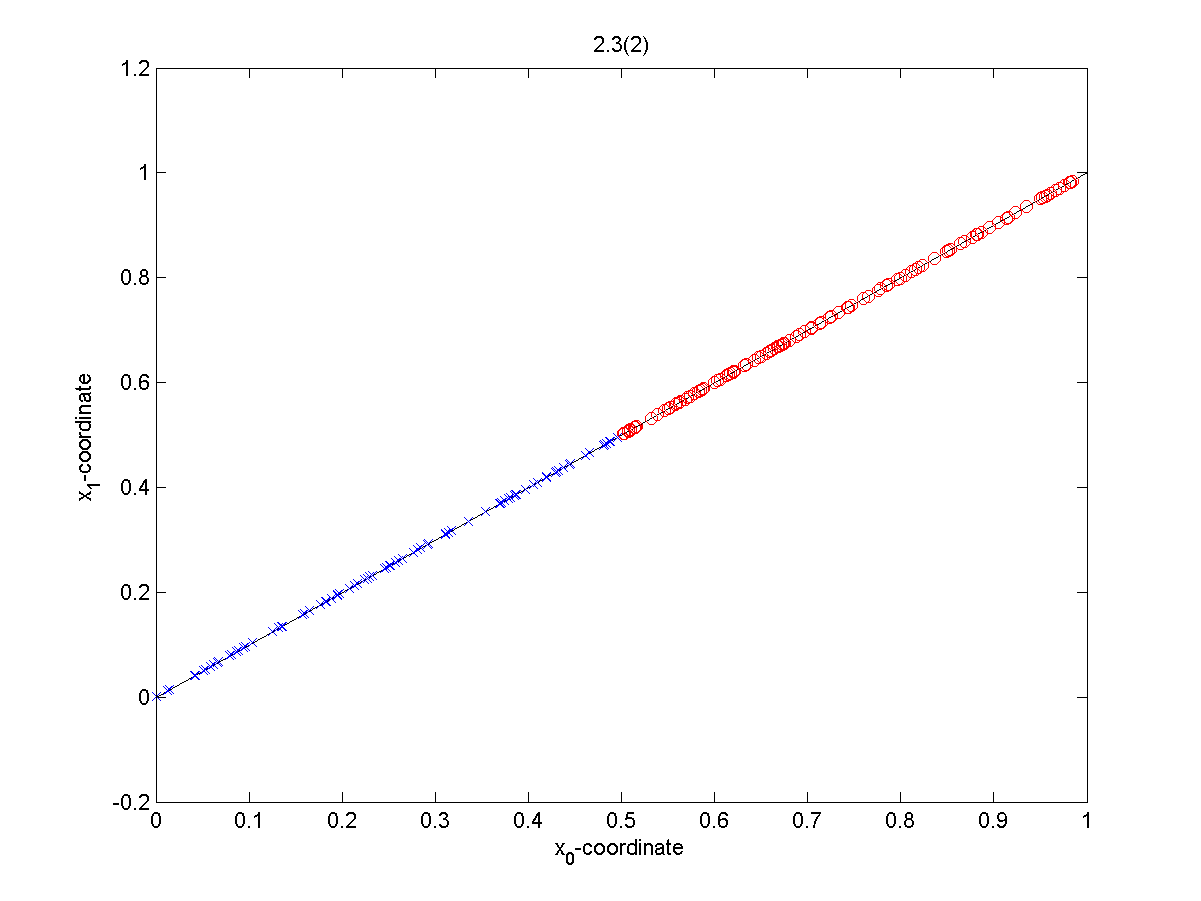


Figure 2. A least-square regression fail to separate the data

Although we get a perfect least-squares linear regression, it is fail to separate different symbol. The situation occur at the different symbols are almost lie on one line, the regression line.

* 1. We rewrite the formulation in matrix from:

Where, , , .

We want to get the minimum, find the gradient by:

By simple transpose, we can find out

* 1. Plot the figure below:

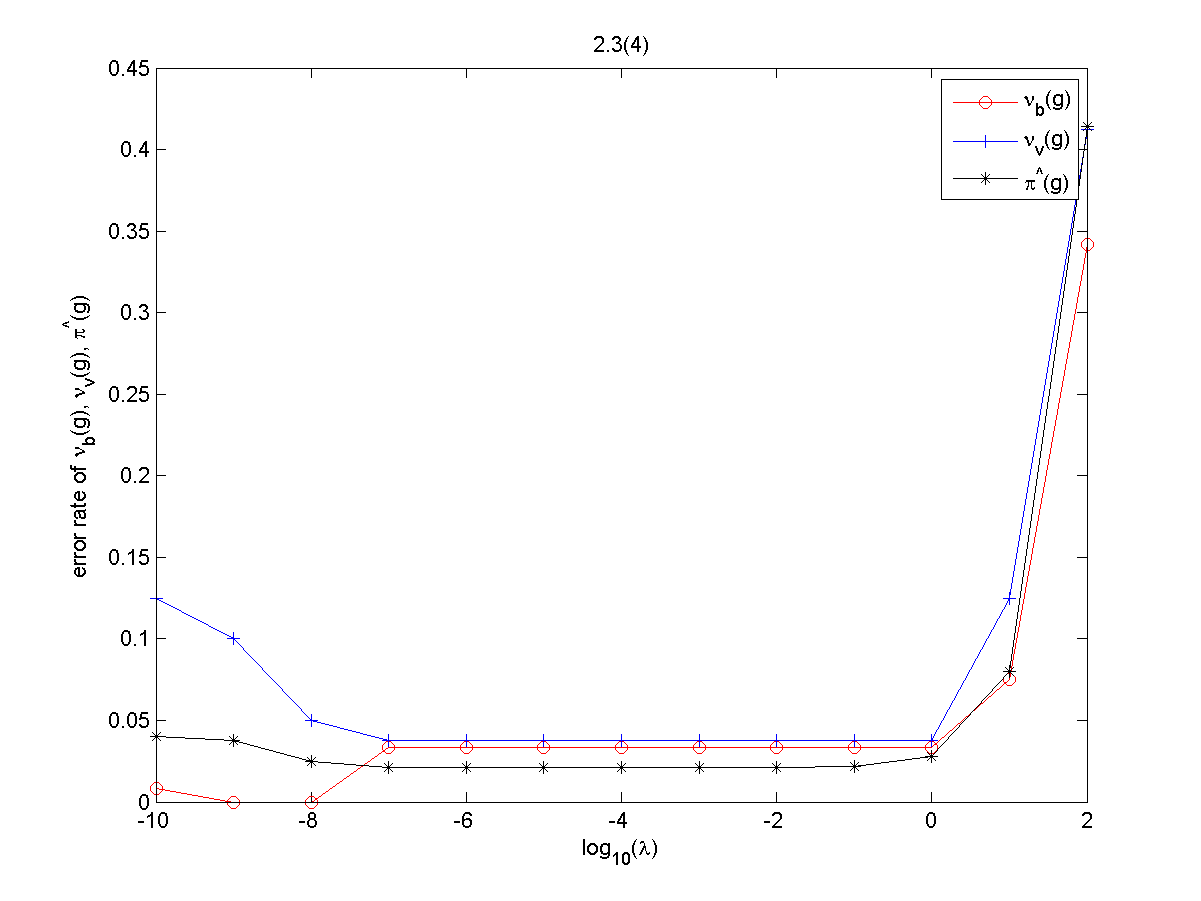


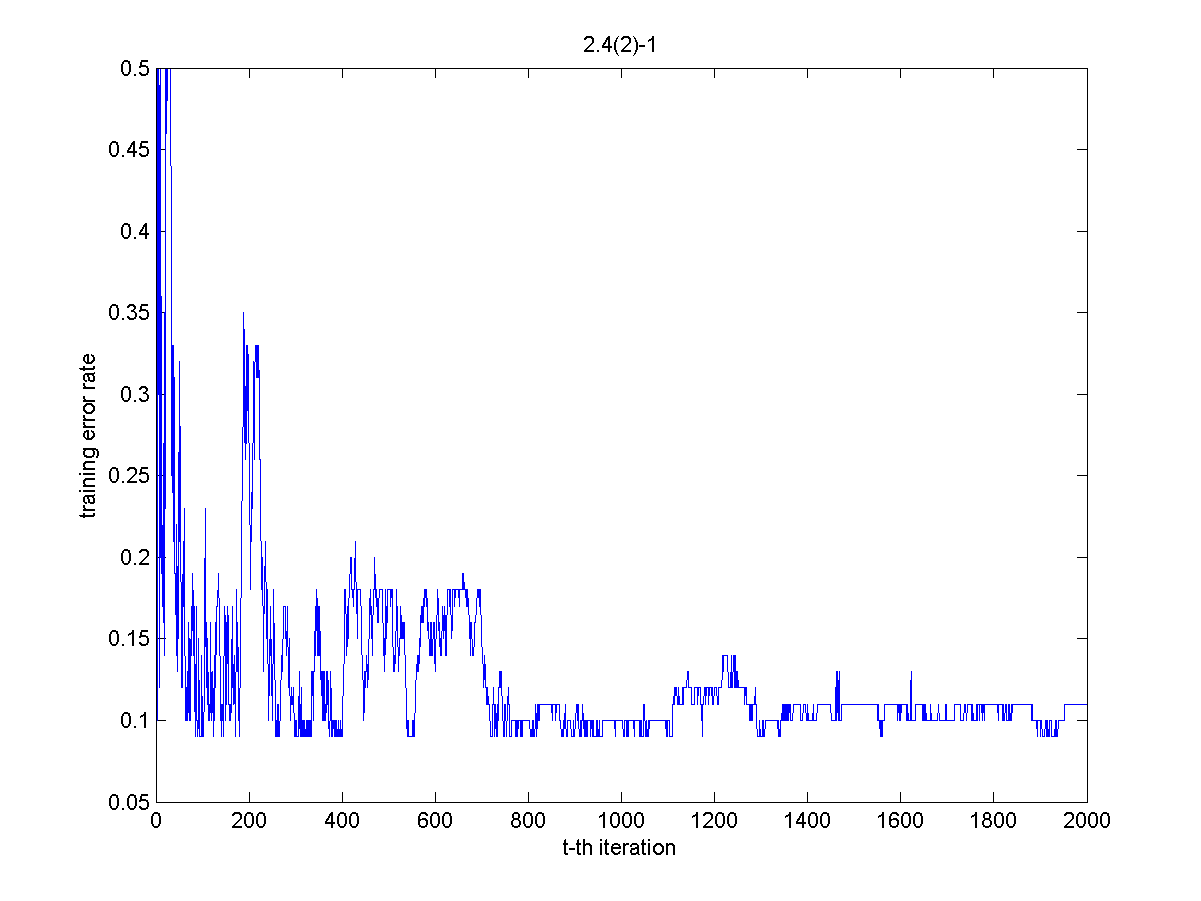
Figure 3. Error rates with different

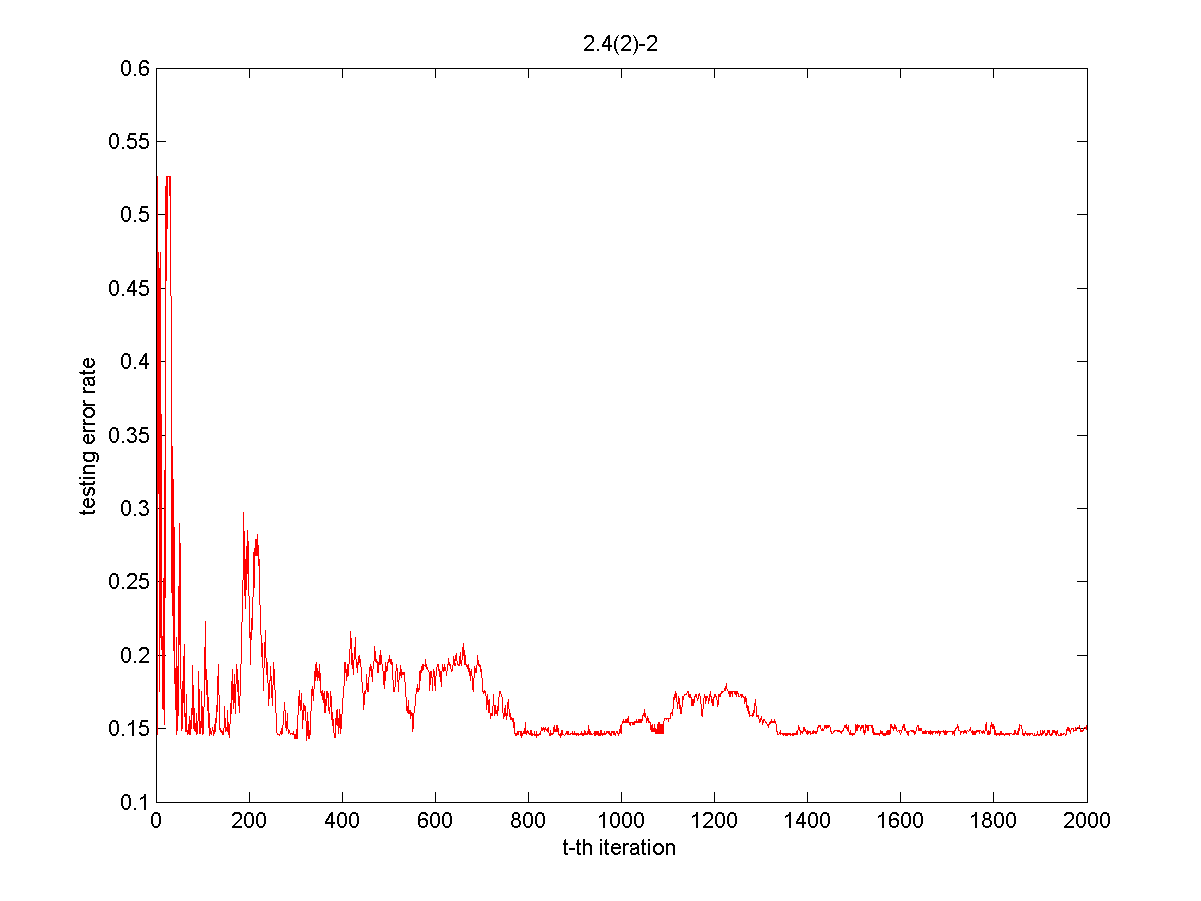
At the beginning (), we find that the is not so close with , is closer. However, when , and are really close, that is what we want. We also can find out that in and aren’t that close, just like the situation at the beginning. When , and are close.

Brief conclusion:

* In some region, and are close, especially at

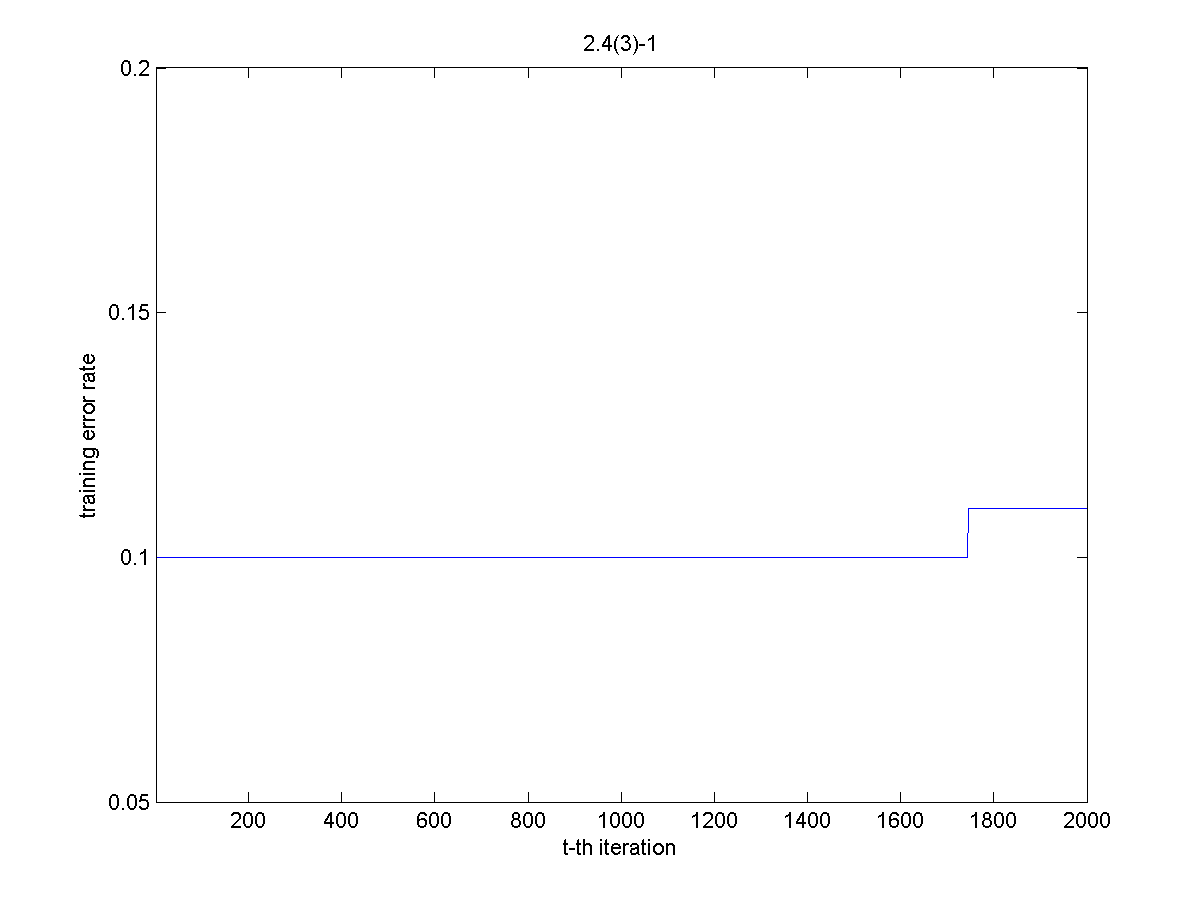
1. **Gradient Decent for Logistic Regression**
   1. We do the partial derivative with on .
   2. The figure show as below:

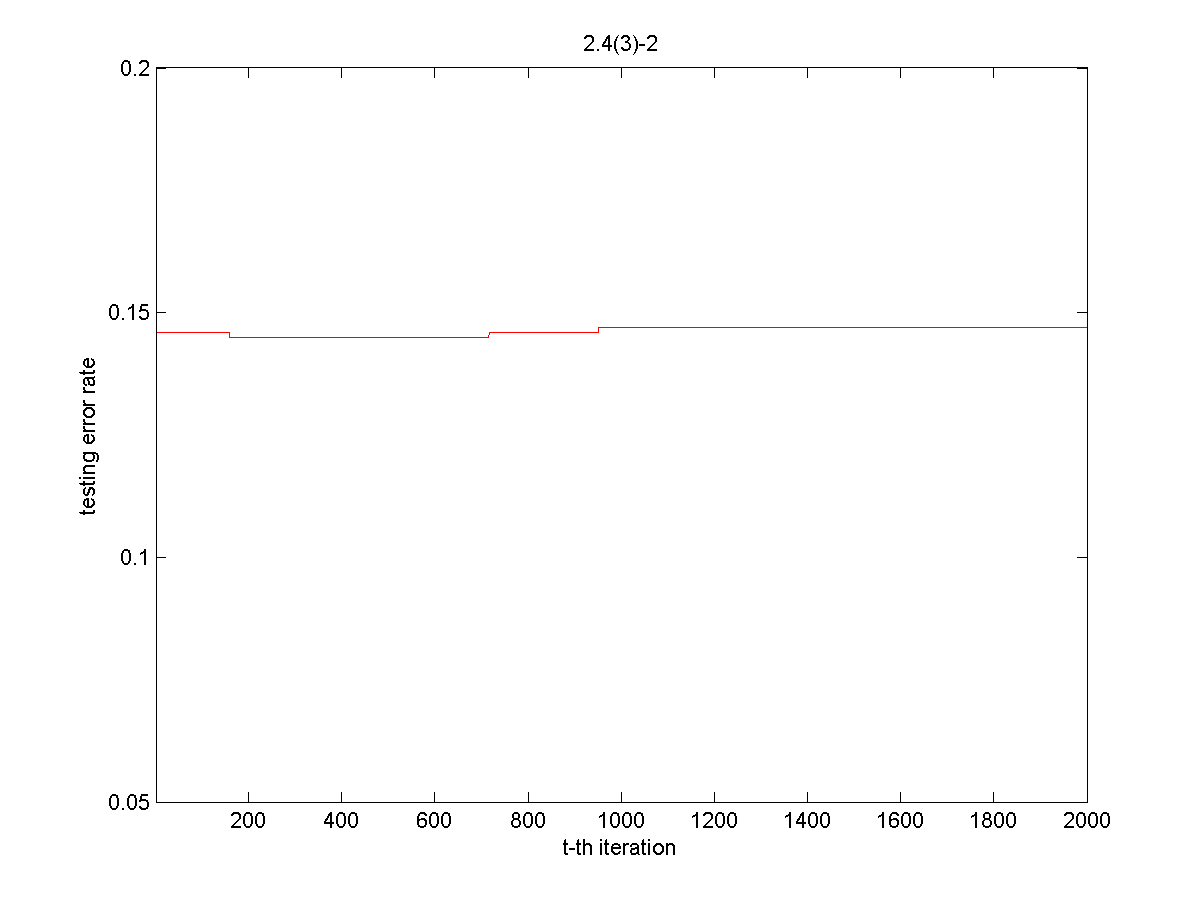




Two figures seem like each other, they both shake at the beginning. But with the t grow up, and it will give our enough training to minimize the E function.

* 1. The figure show as below:





The obivious different between figures of and figures of is that is better than at beginning and it can hold this good sperated line at the end, because we condsider about all training data at every step. However, we find out the error may increased when t get bigger, it may imply the result that we have over training the decision function.