**Homework #4**

1. **More about VC Dimension**
   1. We apply an exponentiation to both side of the inequality

Which mean the number of all patterns we may generate is less or equal to total number of classifier we have. It is satisfy by the truth of the relation between classifiers and patterns are one to one mapping.

* 1. We apply an exponentiation to both side of the first part of inequality

It is trivial satisfy by the fact of we can at least generate one pattern if we have a classifier, say G’. And we see the second part of the inequality

G’ is the intersection of , It is satisfy by the set theory, the maximum of is exactly set, depend on which one is smaller.

* 1. First part of the inequality also use the set theory, we proof that the first part is satisfy by the fact the minimum of , where, G’’ is the union of , is equal to set, depend on which one is bigger. Therefore,

For second part, assume that the upper bound is . We can separate the bits in two parts, one has bits, and another one has bits. If using G1, the patterns is less than equal to . In the other hand, if using G2, the patterns is less than equal to . The fact maximum numbers of pattern is

Which is smaller than .We find out that each part cannot generate all pattern because the constraint of VCD. Because the , the upper bound will shrink to .

1. **Curse of Dimensionality**
   1. Consider about the volume of hypercube with d dimension can be written by

A hypersphere of radius R in d-dimensions can be represent by a d-tuples of points such that

Let denote the `content’ of a d-hypersphere or radius R is given by

By integration in n-d spherical coordinates (Stewart, 2006), we have a recurrence relation:

Thus, by define and the log ratios for d = 1, 2 … 10 can be showed by figure 4.2(1)

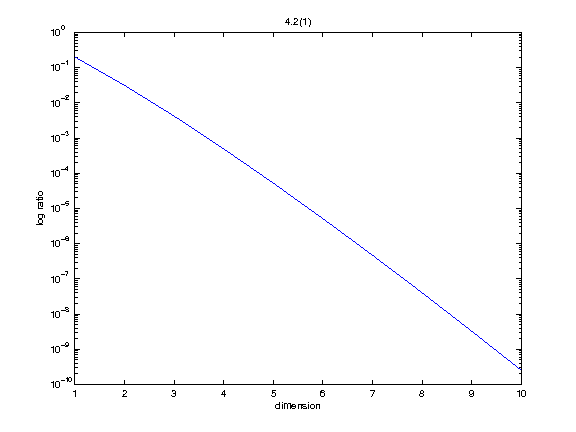


Table 4.2(1)

Figure 4.2(1)

And the ratios are show in Table 4.2(1)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

* 1. We can reduce this problem to problem 4.2(1). Because we generate the points uniformly in the hypercube. The probability of origin’s nearest neighbor in the training set within 0.1 can approach to the ratio we compute above. Hence we want to ensure the probability , where the probability

Therefore, by inequality

We show n in Table 4.2(2),

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Table 4.2(2)

* 1. We can find out the volumes of d-hypersphere inside the hypercube is decrease by exponentiation with base 2, when we move it to a corner of hypercube. We have the new statement

Therefore, by inequality

We have Table 4.2(3)

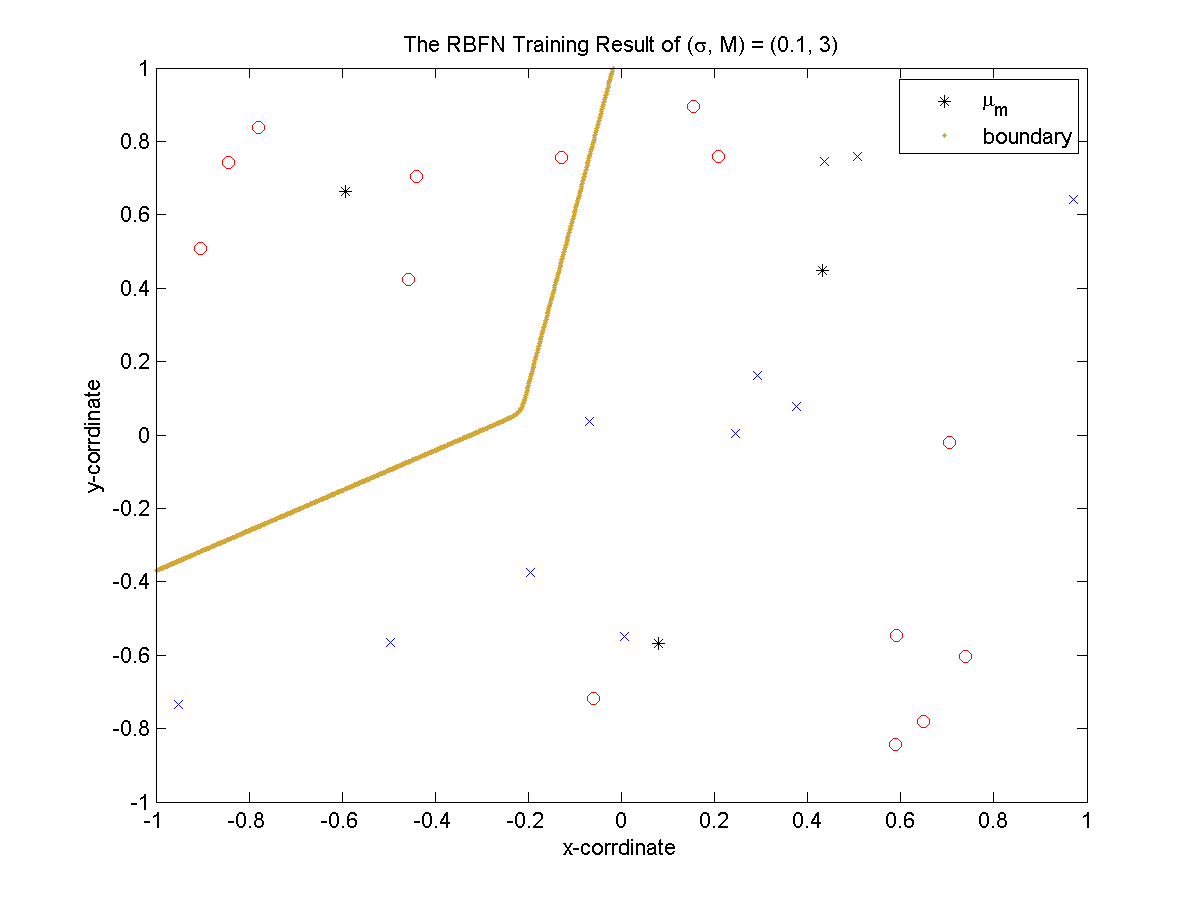
|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

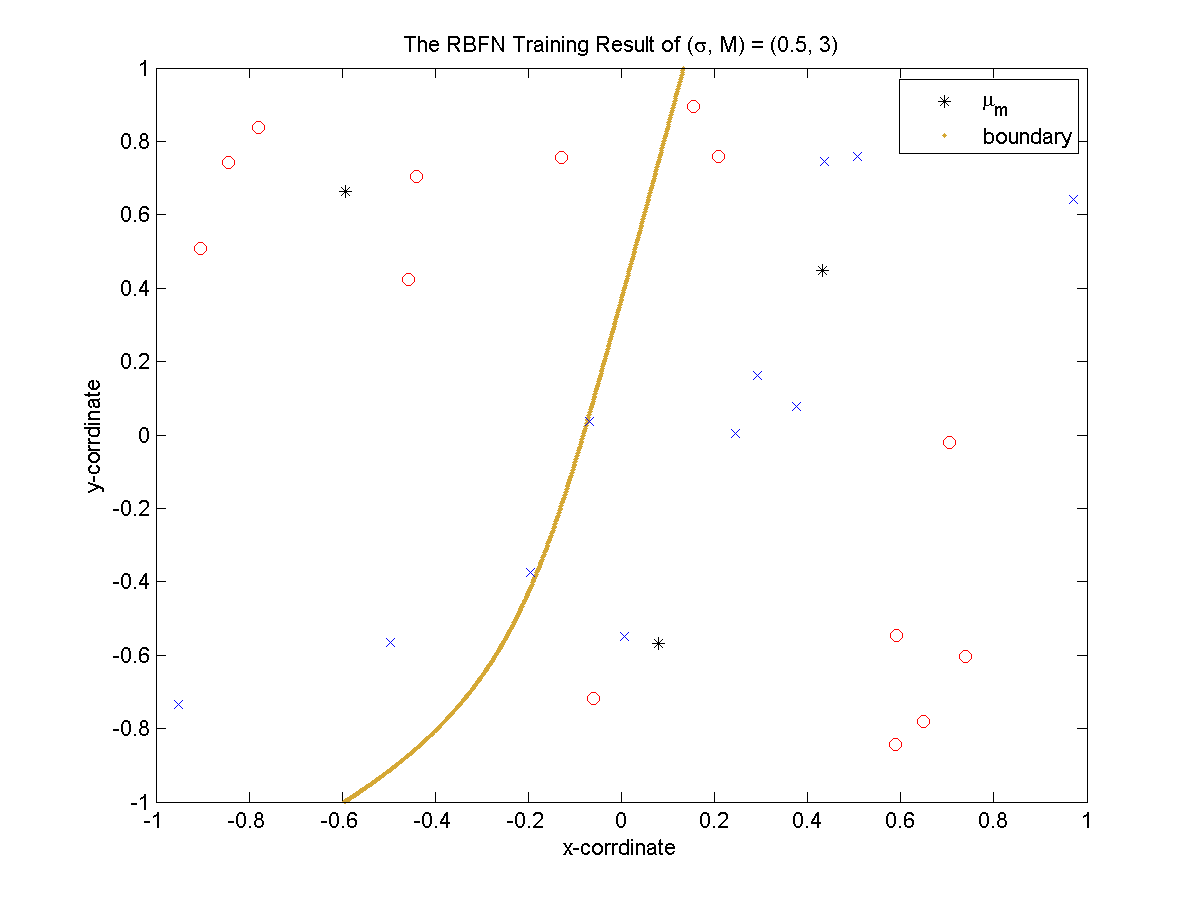
Table 4.2(3)

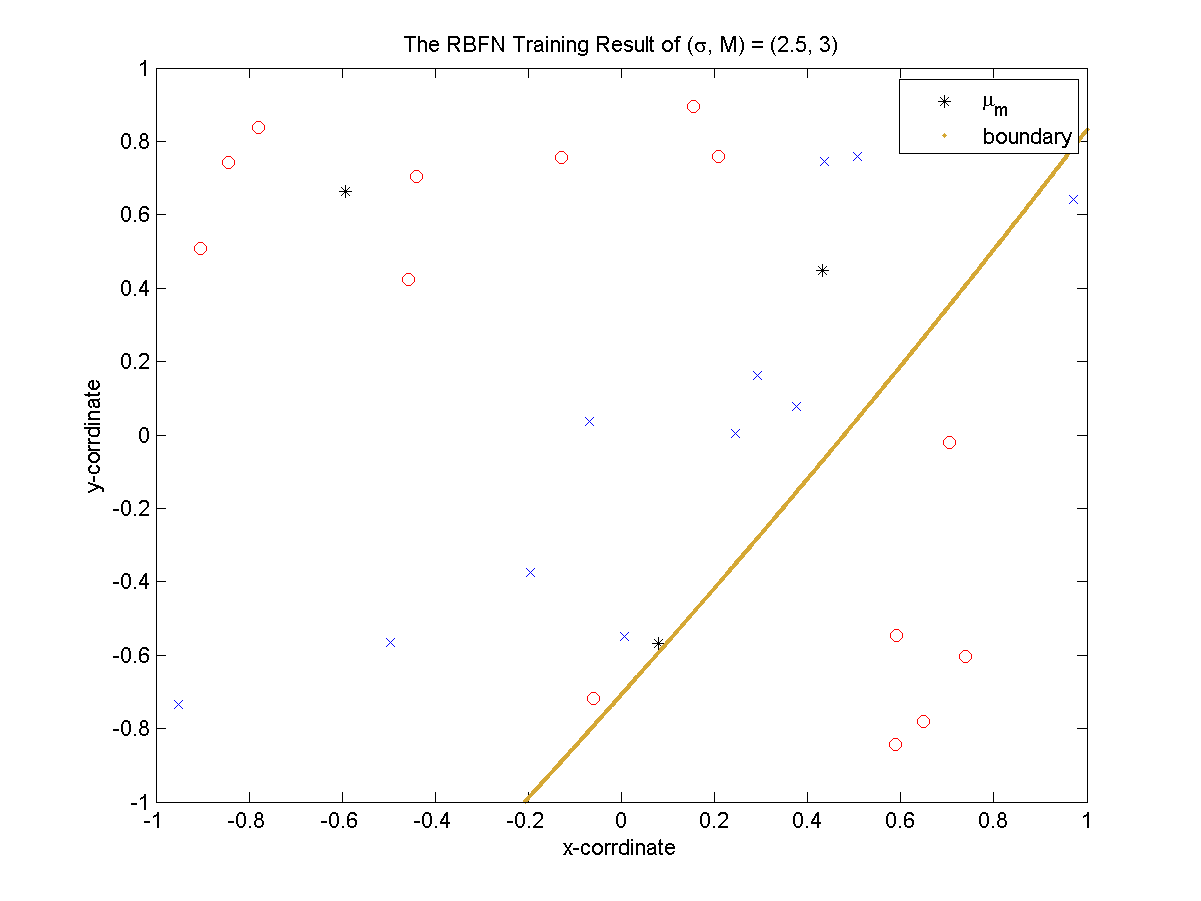
The finding is we may suffer by the dimension of hypercube. For example, we need lots of sample to ensure the data similarity is consisted. When we try to query the point in the corner of testing data, we need exponentiation time sample to achieve that goal.

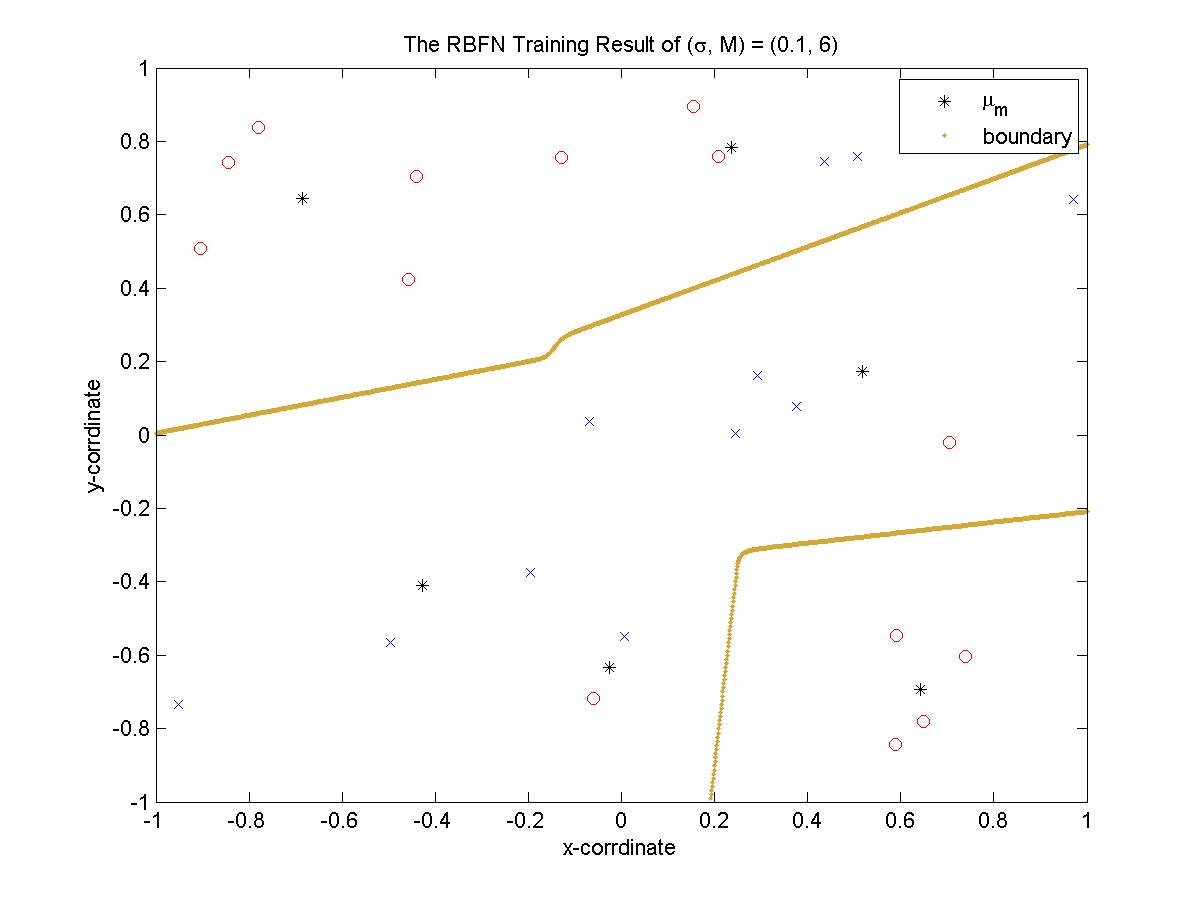
1. **Experiment with Radial Basis Function Network**
   1. Some brief finding
      1. When M too small, we cannot get a good result on Radial Basis Function Network.
      2. When delta value gets bigger, we will get a smooth boundary.

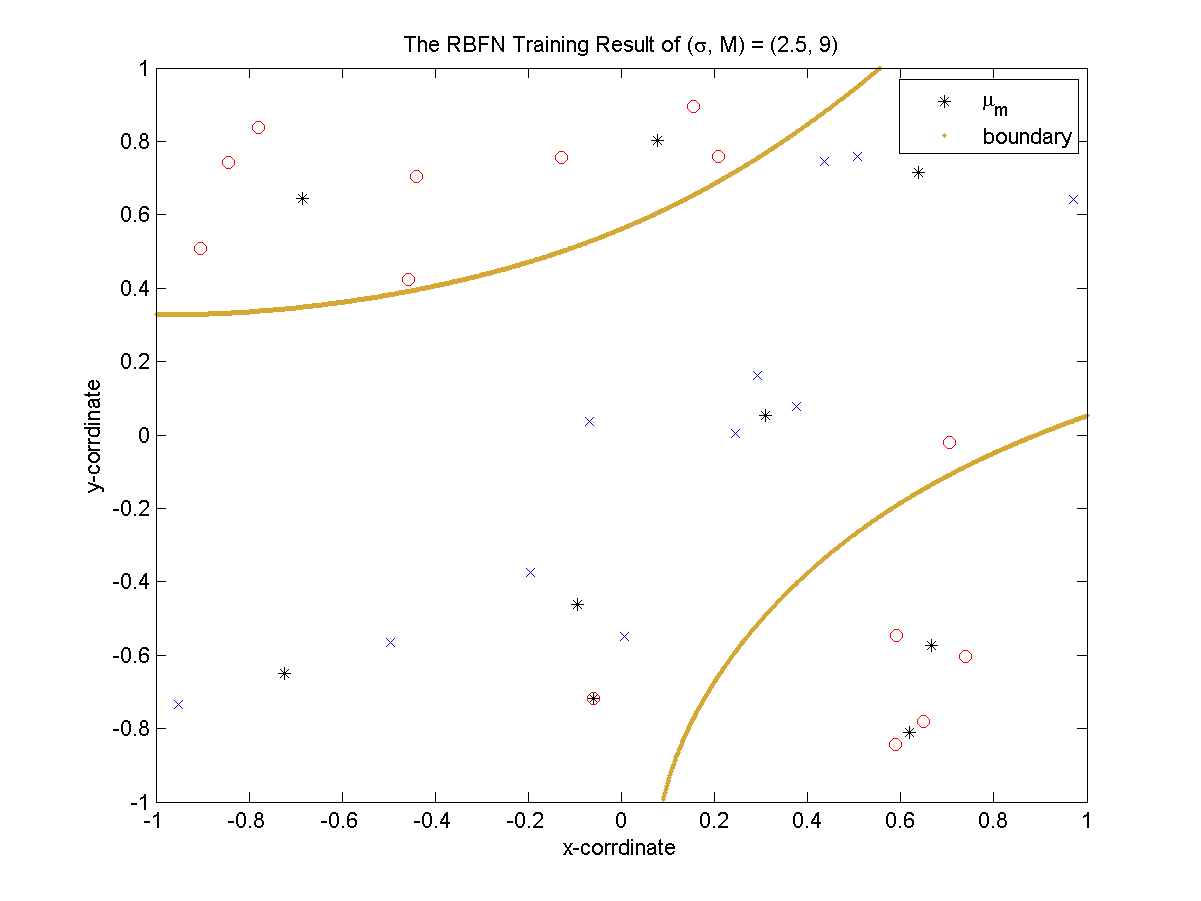
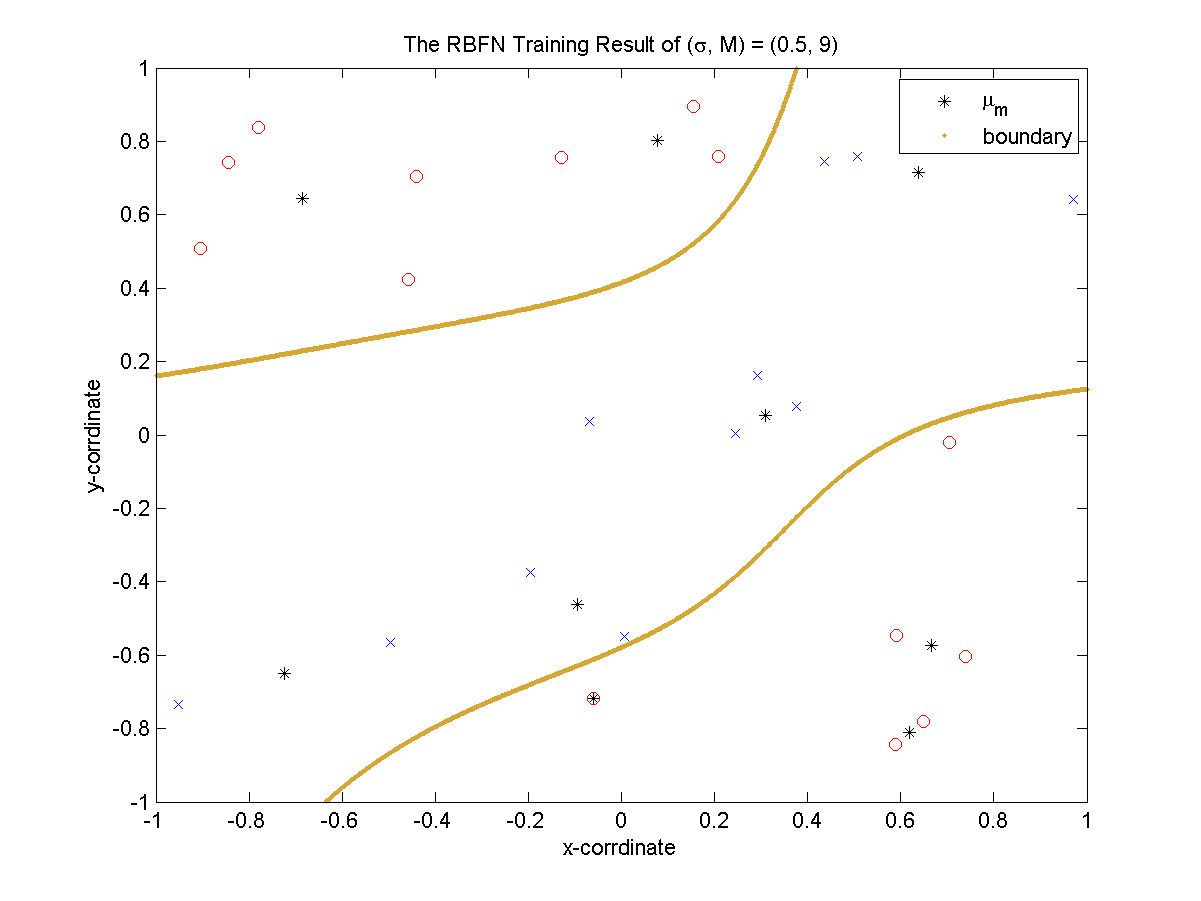
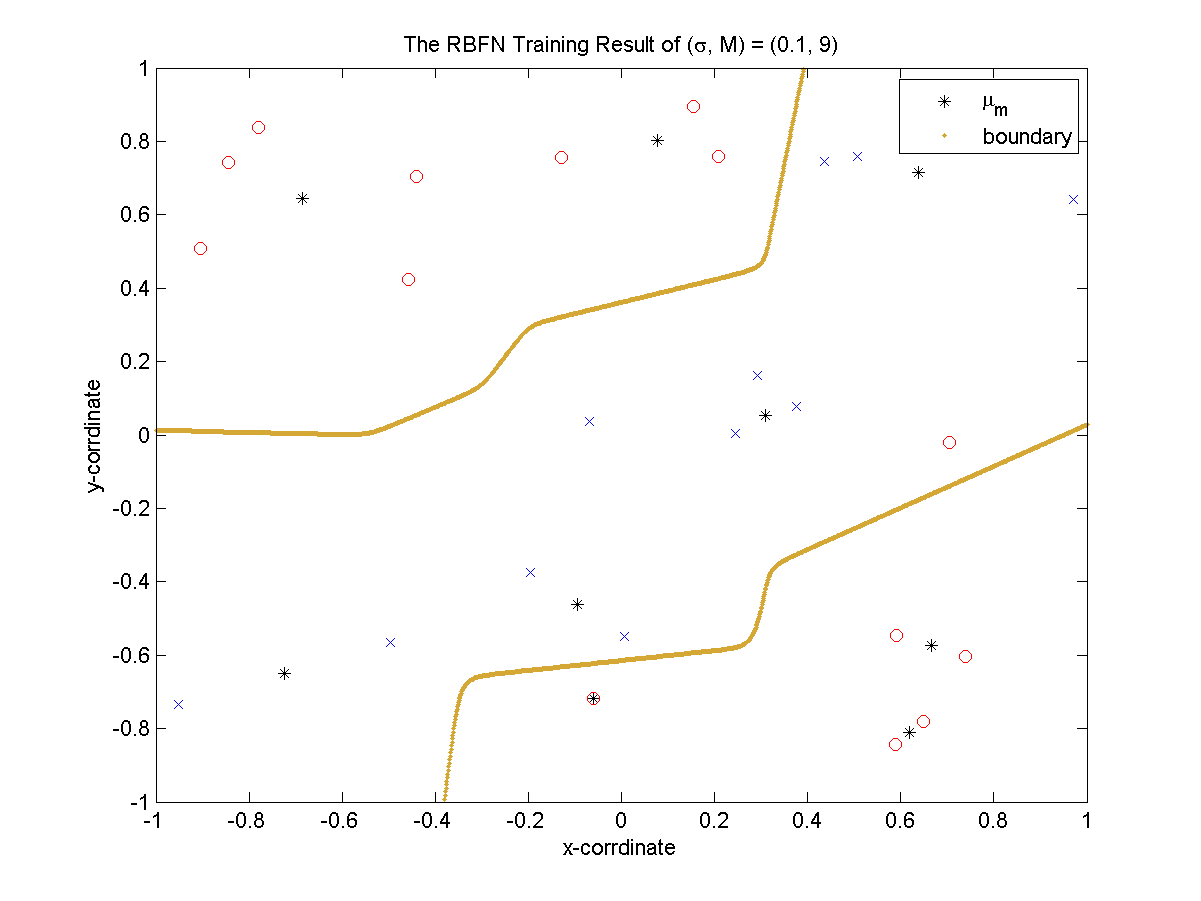
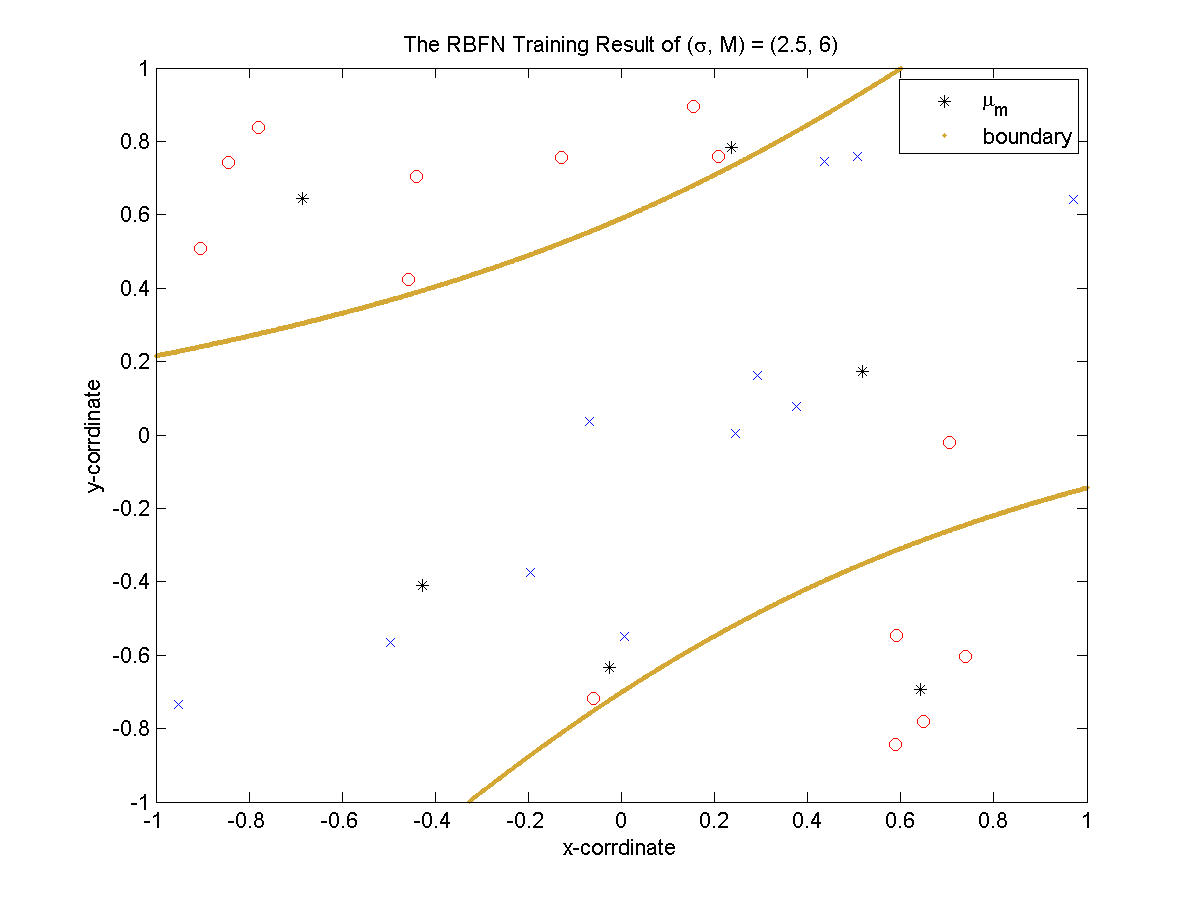
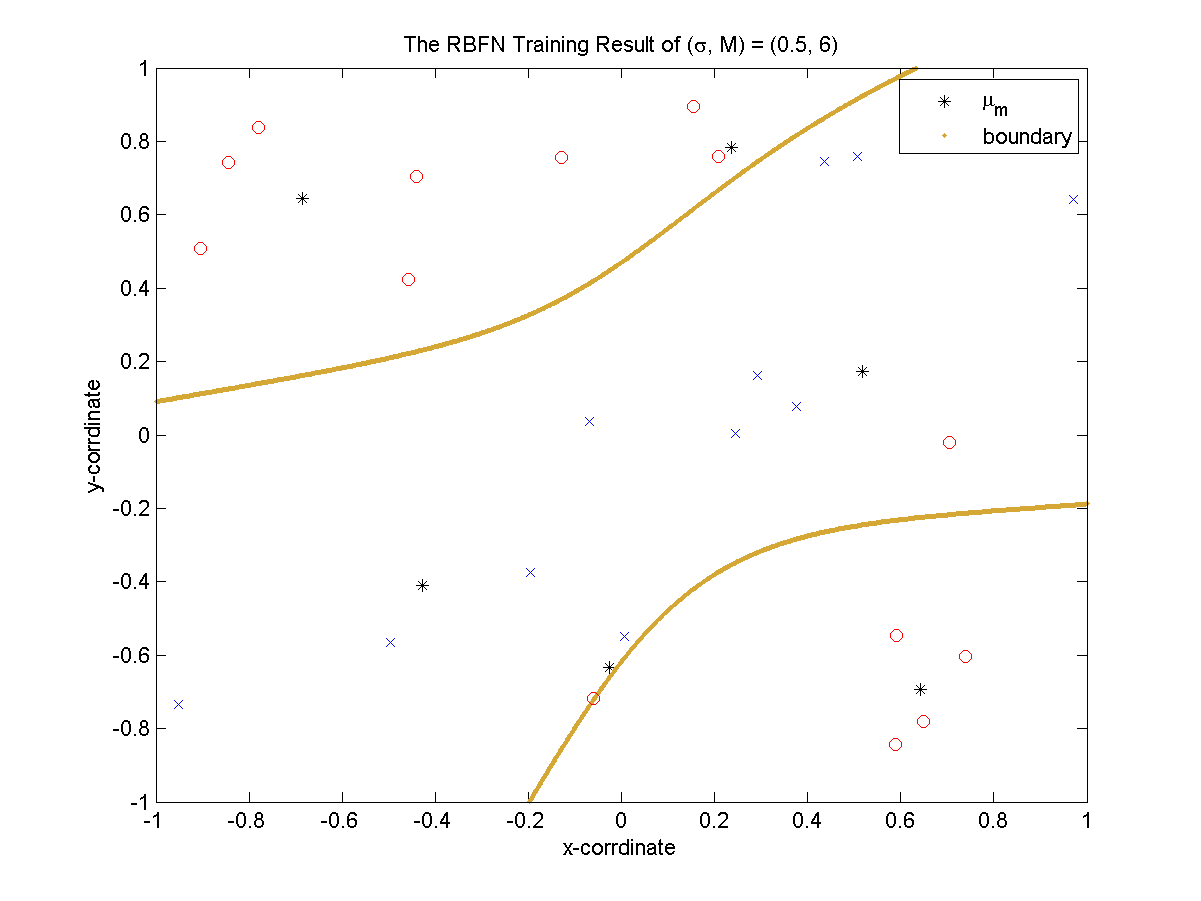
The figures show below











* 1. The values show below

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

We get the best result on with , which is not have a best training error. The fact imply one thing, those get lower training error one may over fitting the training data, like with .

1. **Experiment with Backprop Neural Network**
   1. The conclusion is
      1. With a large learning rate, the E value cannot converge in a good place, which means we get a bad result in every case with big learning rate.
      2. With a proper M, say 3, we can get a good result with a lower E value.
      3. The random initial value looks it not that important in our case.

The figures show as below:

