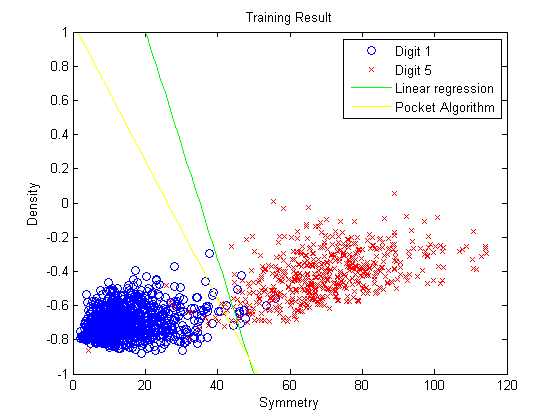
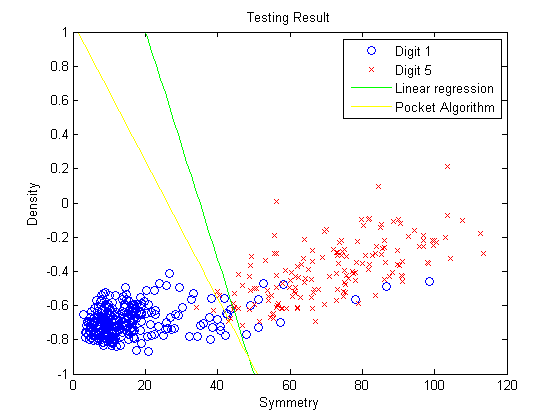
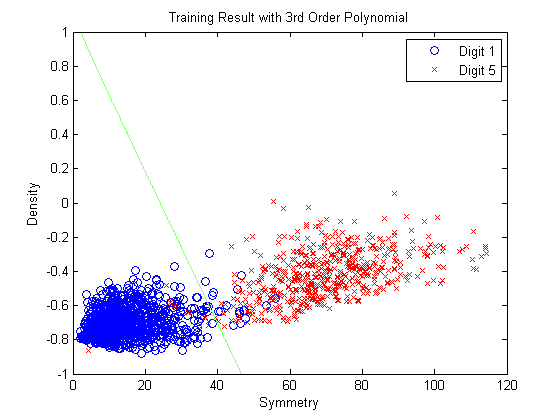
1. **Digits Classification**
2. I would use Linear Regression for classification followed by pocket for improvement.

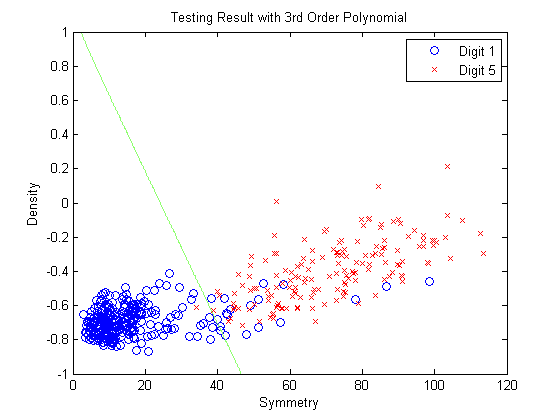
The plots for training and testing data with linear classifiers is as below:





1. The bound for is
2. Using a third Polynomial linear classifier, the classification result is as below:

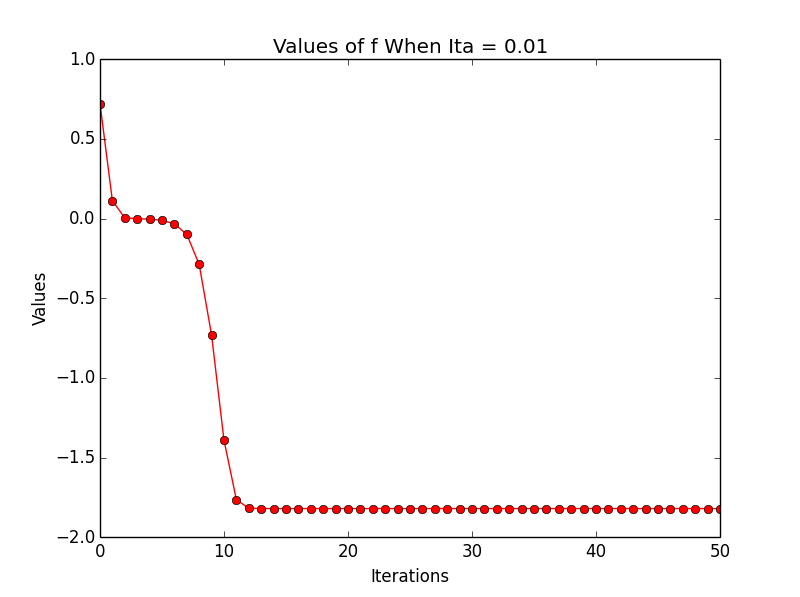




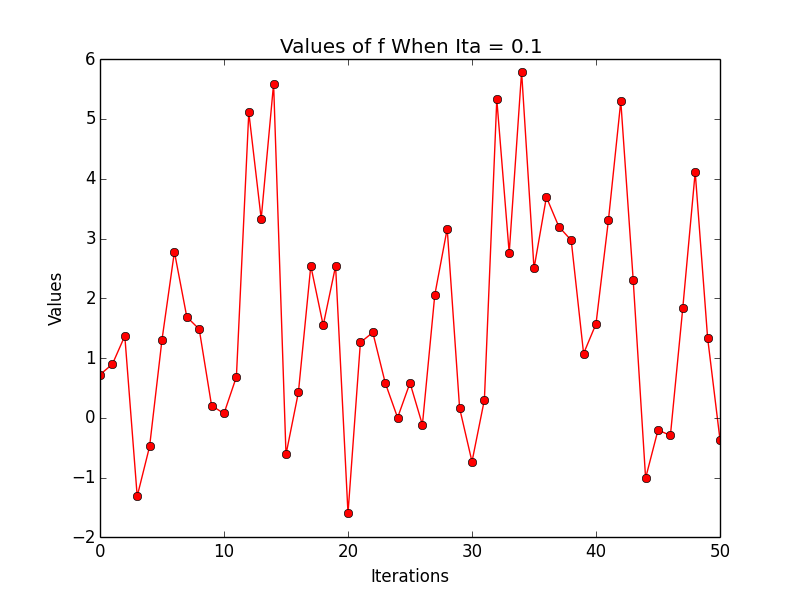
Where .

**2. Gradient Descend**

(a) When , the function value in different iterations are as below:

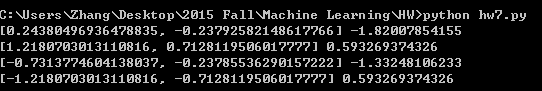


When , the function value in different iterations are as below:



It can be seen that the optimal function values converge fast (less than 15 iterations) when is relatively small; but face the problem of convergence when is large. Thus when the learning rate it too large, the GD may not converge. An appropriate choice of is needed.

(b)



The results with different initial points are as below:

|  |  |  |
| --- | --- | --- |
| Initial Points | Sub-optimal Points | Sub-optimal Value |
| (0.1,0.1) | (0.2438,-0.2379) | -1.8201 |
| (1,1) | (1.2181,0.7128) | 0.5933 |
| (-0.5,-0.5) | (-0.7313,-0.2379) | -1.3325 |
| (-1,-1) | (-1.2181,-0.7128) | 0.5933 |

It can be seen that different initial points give us different minimum solutions. These solutions are local minimum and difficult to improve by the GD algorithm, so it is difficult to find global optimum solutions.