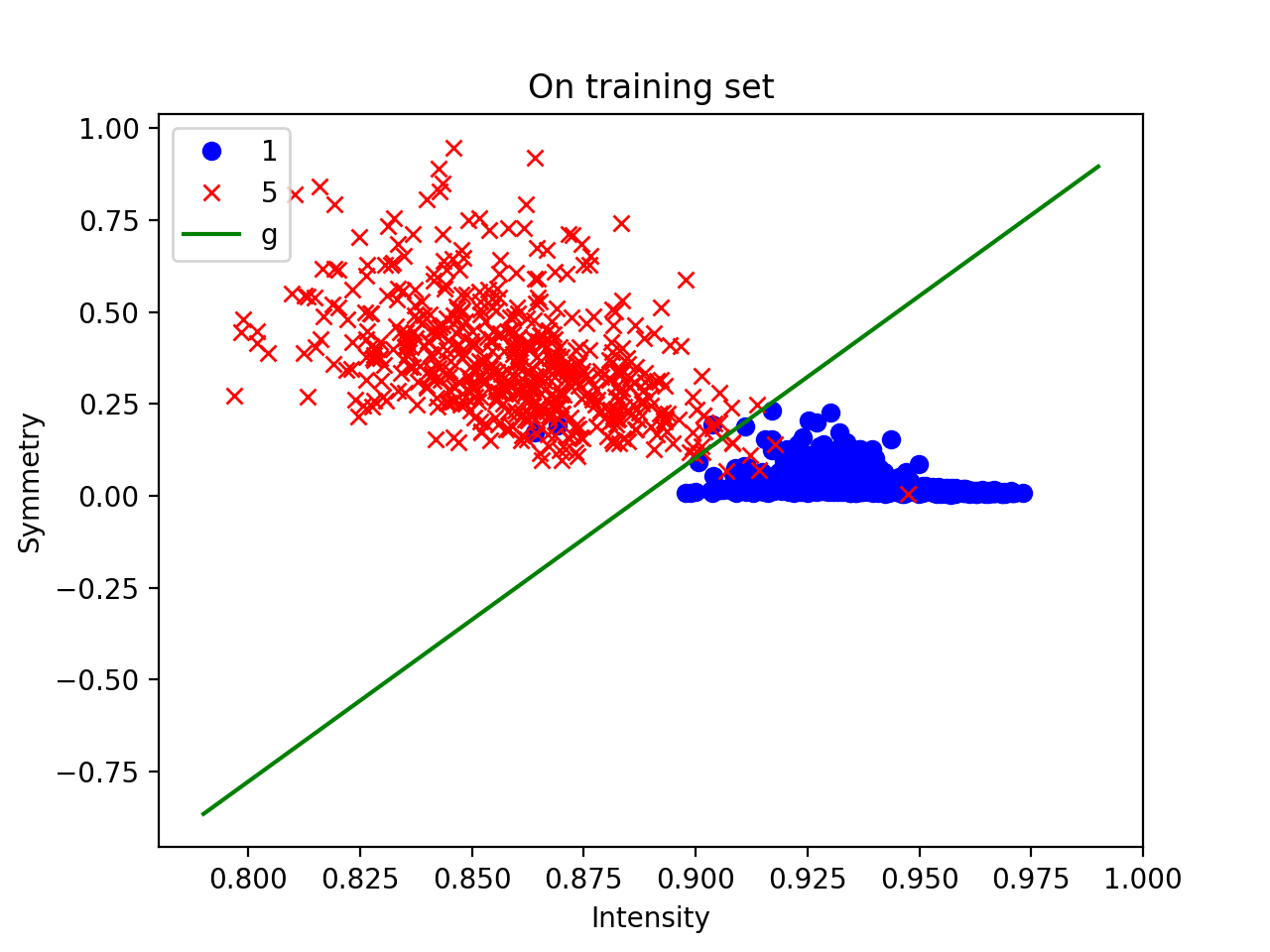
CSCI 4100 Assignment 7

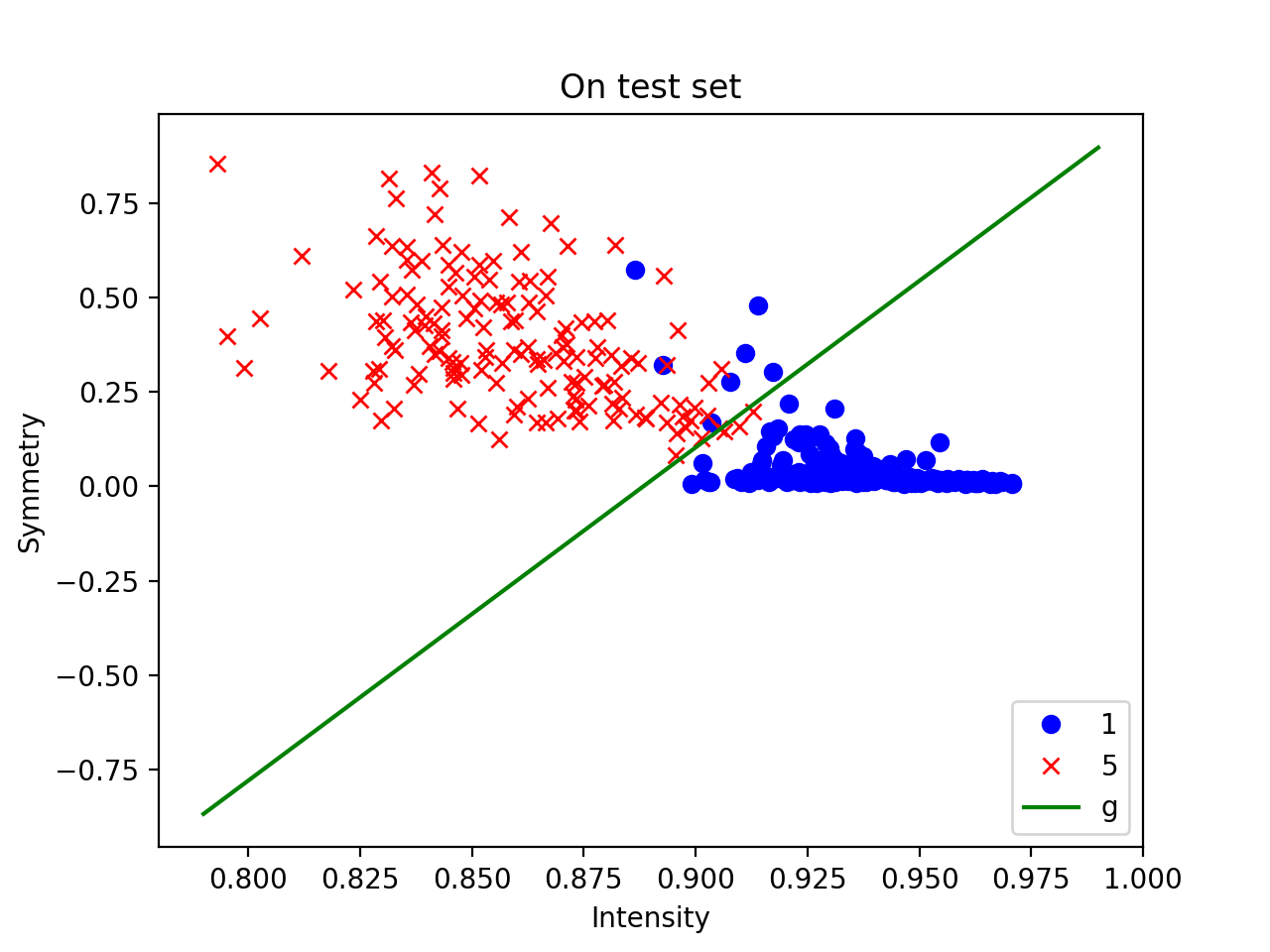
Boliang Yang 661541863

1. Classifying Handwritten Digits: 1 vs. 5

I pick Linear Regression for classification followed by pocket for improvement.

1. With , we compute initial weight and then update with pocket algorithm. The final hypothesis is

Plot of the training data:

Plot of the test data:

1. ,
2. Error bound based on :

Since for a linear perceptron in two dimensions, . Since , then

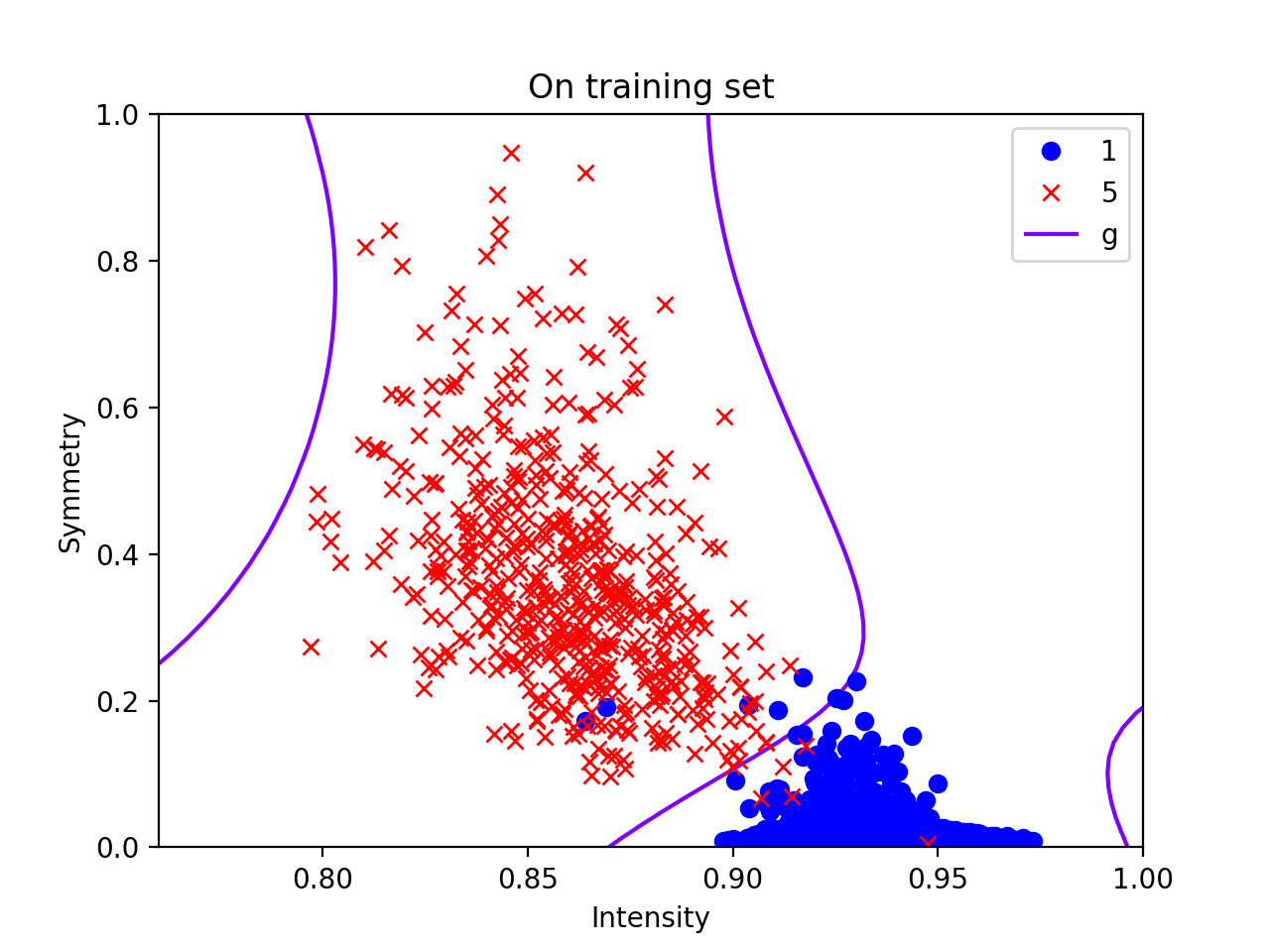
Error bound based on :

Since there is only one hypothesis, we can simply use Hoeffding bound for the error bar from the test data set:

Since , we have , then .

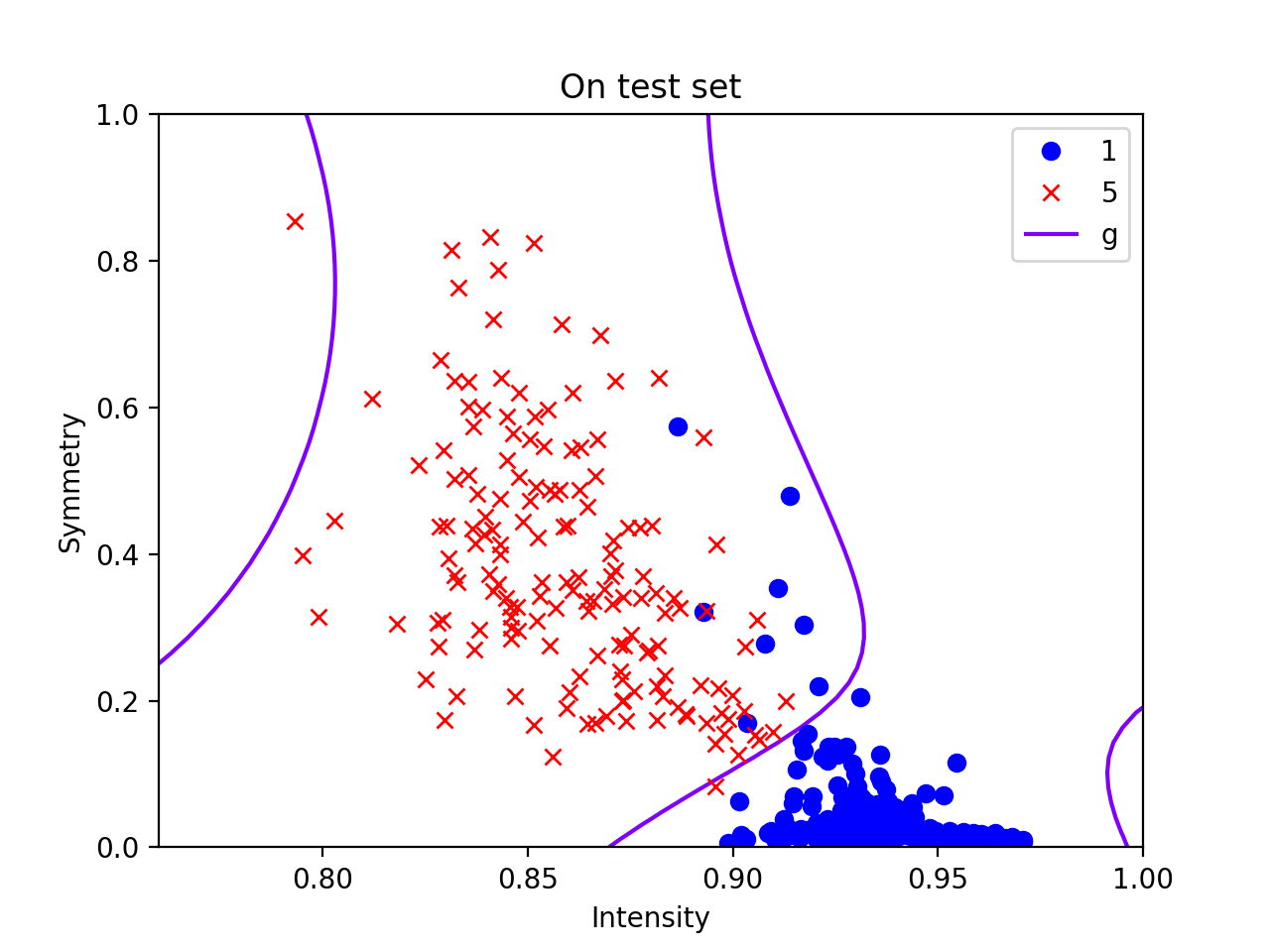
Thus, the error bound based on is the better bound, and this is because that it has only one hypothesis rather than all linear hypotheses.

1. Now we use a 3rd order polynomial transform, we replace with . The final hypothesis is

Plot of the training data:

Error bound based on :

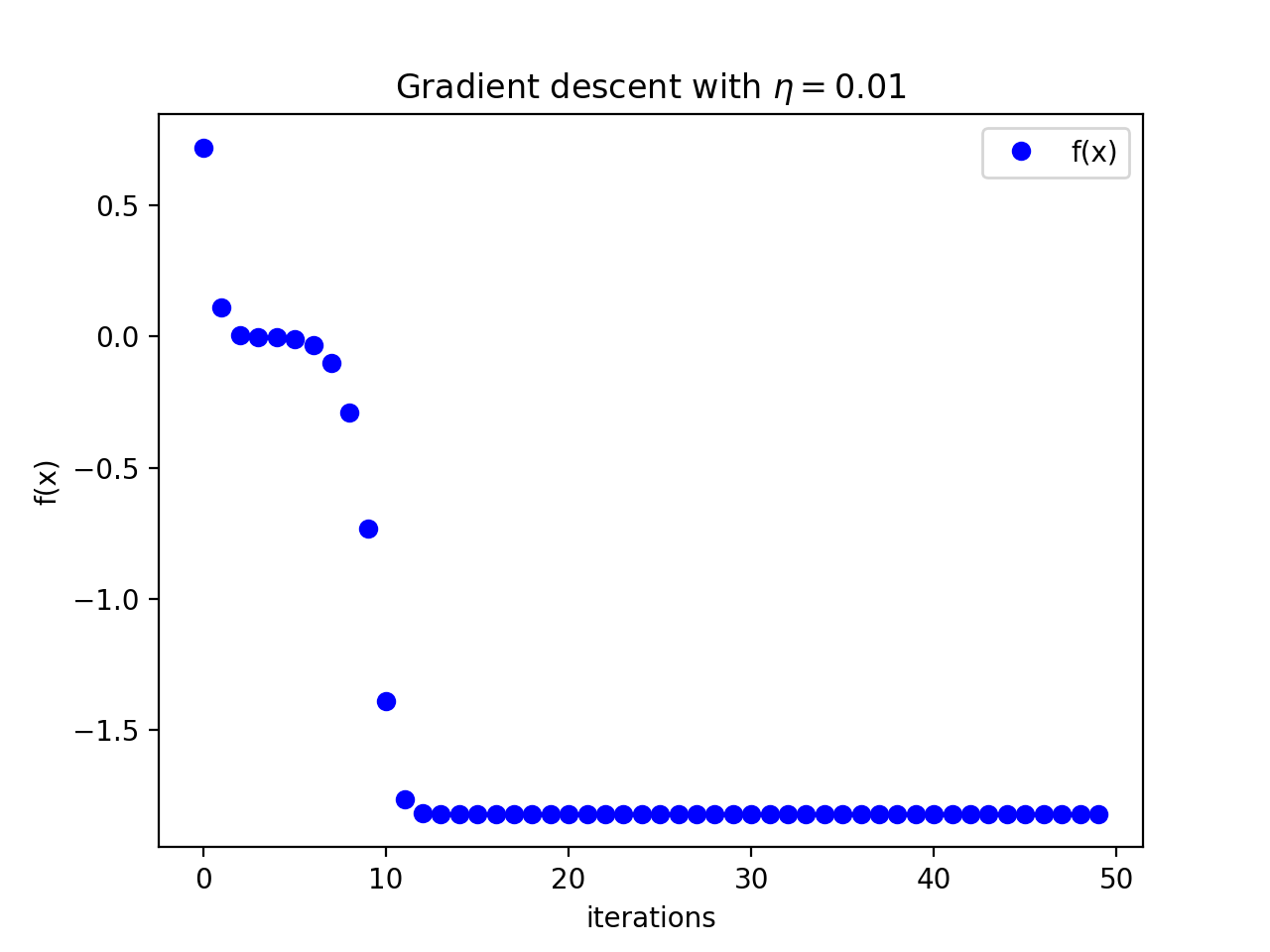
Since from problem we have , , then

Plot of the test data:

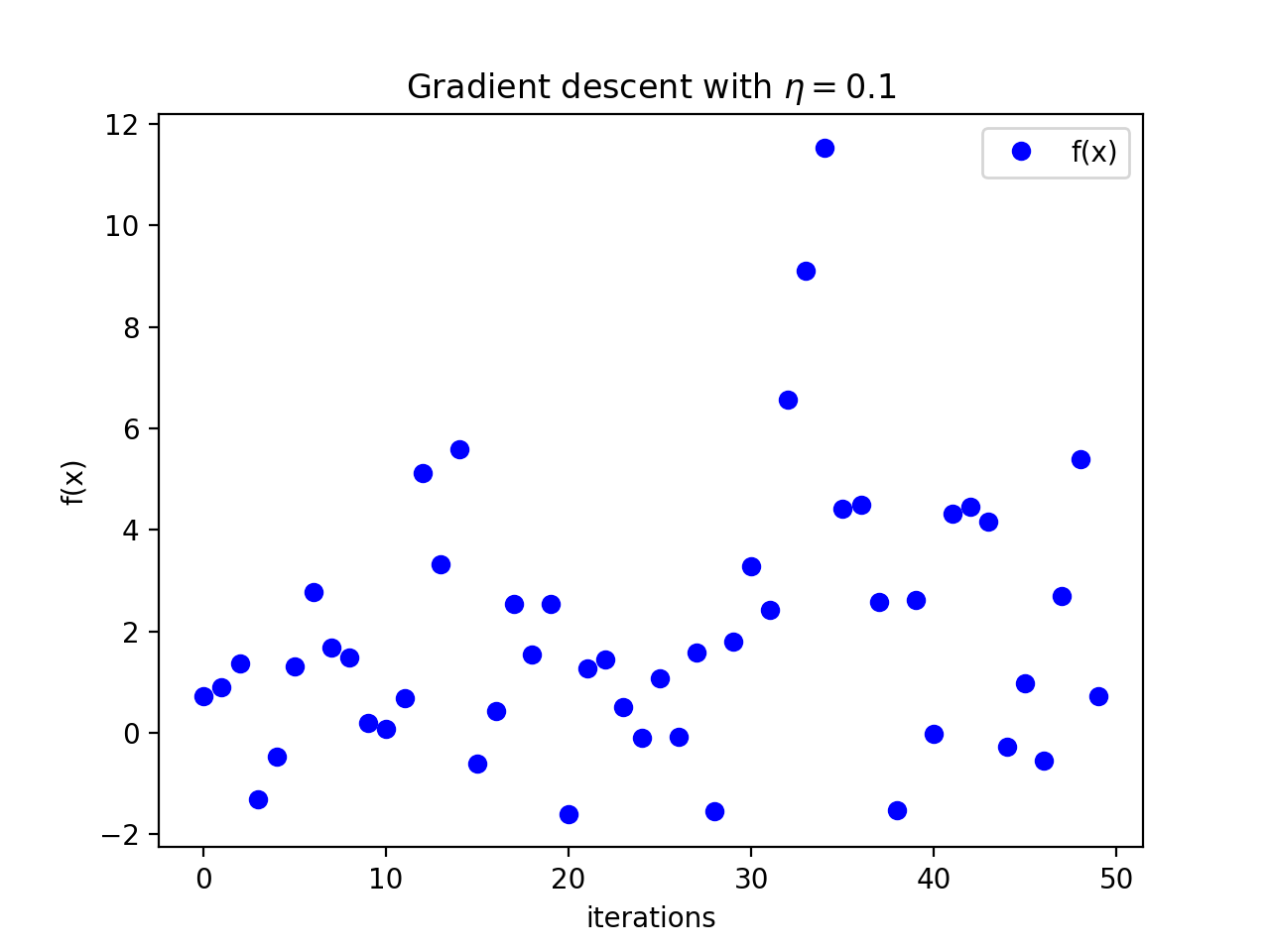
Error bound based on :

Since , we have , then .

Thus, the error bound based on is the better bound, and this is because that it has only one hypothesis rather than all linear hypotheses.

1. The linear model without the 3rd order transform is better. The two models produce the same , then by Occam’s razor principle, we should choose the hypothesis as simple as possible: the simplest model that fits the data is also the most plausible.
2. Gradient Descent on a “Simple” Function
3. The plot of gradient descent with :

The plot of gradient descent with :



With , the algorithm quickly finds the minimum in 50 iterations; with , the minimization cannot occur because gradient descent keeps overstepping the minimum.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Initial point |  | |  | |
| Location | Minimum value | Location | Minimum value |
| (0.1,0.1) | (0.2438,-0.2379) | -1.8201 | (0.2036,-0.1742) | -1.6005 |
| (1,1) | (1.2181,0.7128) | 0.5933 | (-0.6838,-0.1101) | -0.6753 |
| (-0.5,-0.5) | (-0.7314,-0.2379) | -1.3325 | (-0.2547,0.2840) | -1.7276 |
| (-1,-1) | (-1.2181,-0.7128) | 0.5932 | (0.6838,0.1101) | -0.6753 |

Finding the “true” global minimum of an arbitrary function is a hard problem because the output from the gradient descent could change depending on the initial positions and step distance.

1. Problem 3.16 in LFD
2. Since the cost only exist when wrong classification occurs, then ; and
3. Since we accept if , then . And it follows that , which is . Thus,
4. For the Supermarket: then

For CIA: then

Intuition: If is larger and is smaller, then will be larger.