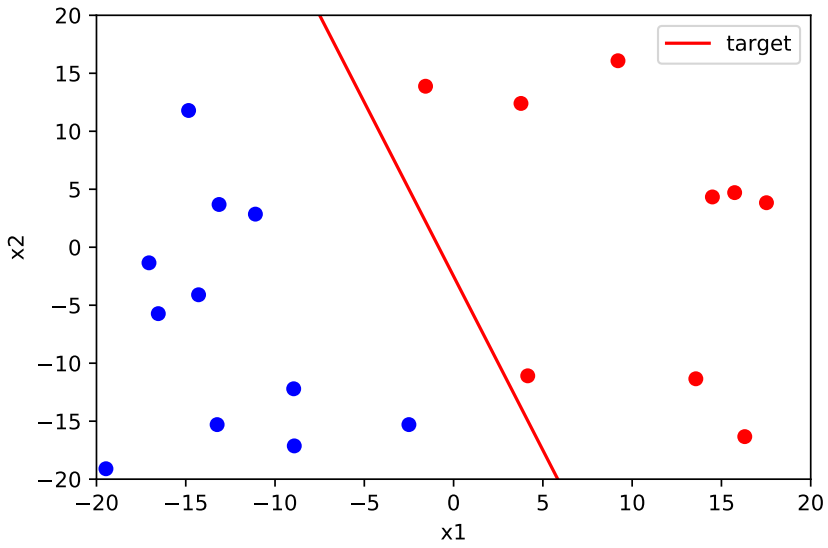
Problem Set 4

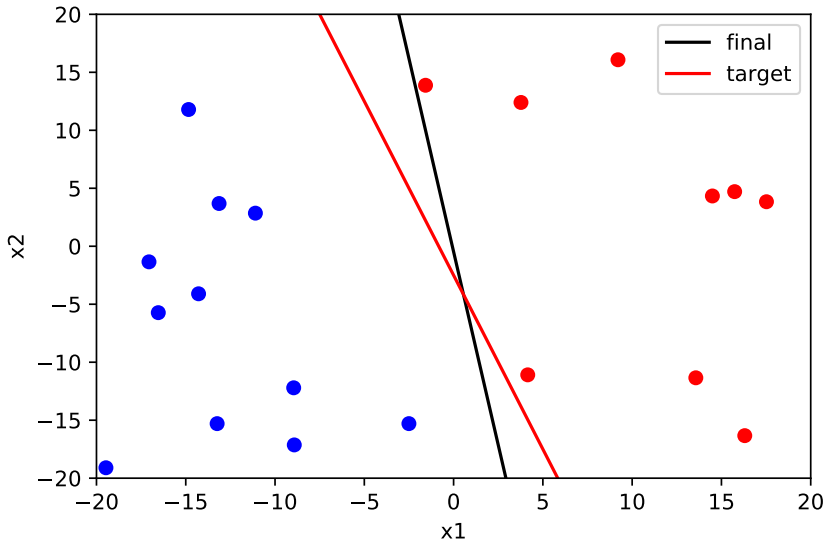
ALM 1.[4,5]; 4.[2]

1. EX 1.4 answer here.

   (a)



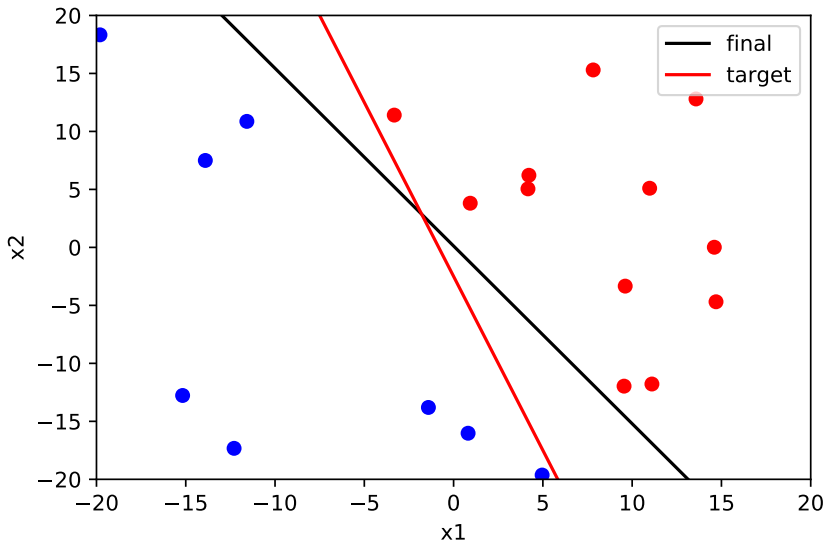
   (b)



   It takes 2 iterations to converge. Even though the final hypothesis, g, correctly classifies data points, g is not very close to f, our target function.

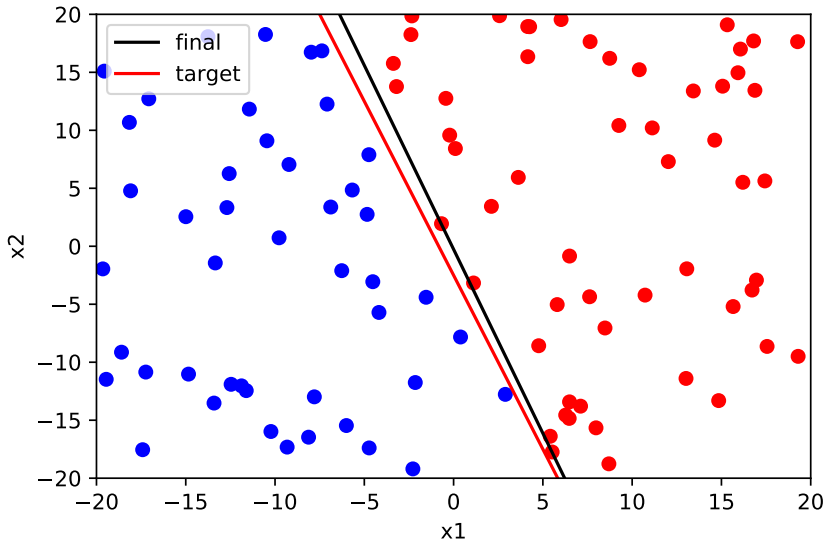
   (c)

Problem Set 4



It takes 5 iterations to converge. The final hypothesis g can correctly classify data points. However, it is not very close to target function, f. g got in this question is even less close to f than g got in (b).

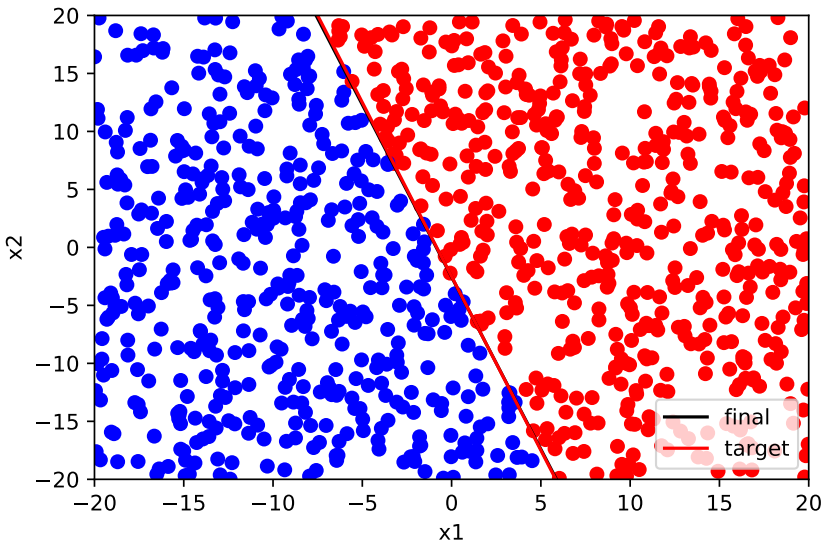
(d)



It takes 38 iterations to converge. The final hypothesis g is pretty close to the target function f. It is far closer to f than g in (b), but it takes more iterations.

(e)

Problem Set 4

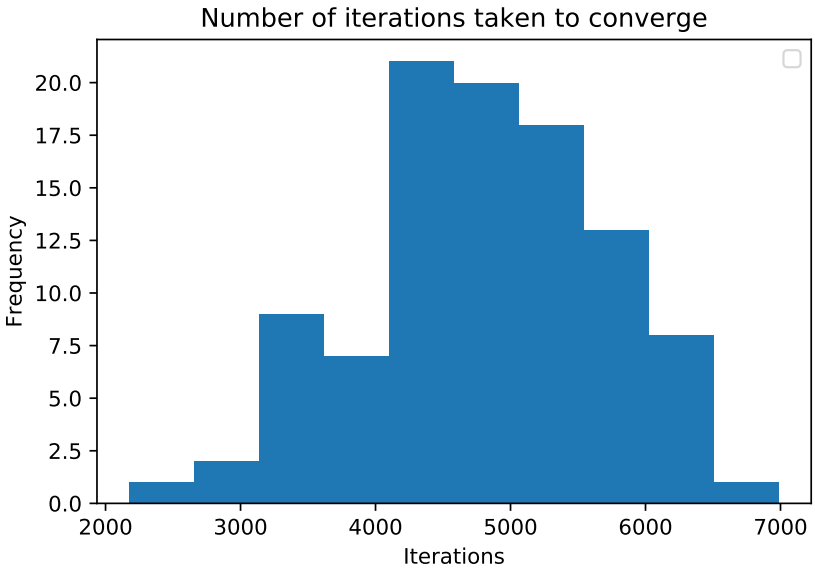


It takes 244 iterations to converge. The final hypothesis g is super close (almost identical) to the target function f. It is far closer to f than g in (b), but it takes far more iterations.

(f)

It takes 5642 iterations to converge, which is pretty big. It is due to the increase of dimension of input (from 2 to 10).

(g)



Problem Set 4

From the 100 experiments, I get the distribution of number of iterations that looks like a normal distribution that is ranged from 2177 to 6987.

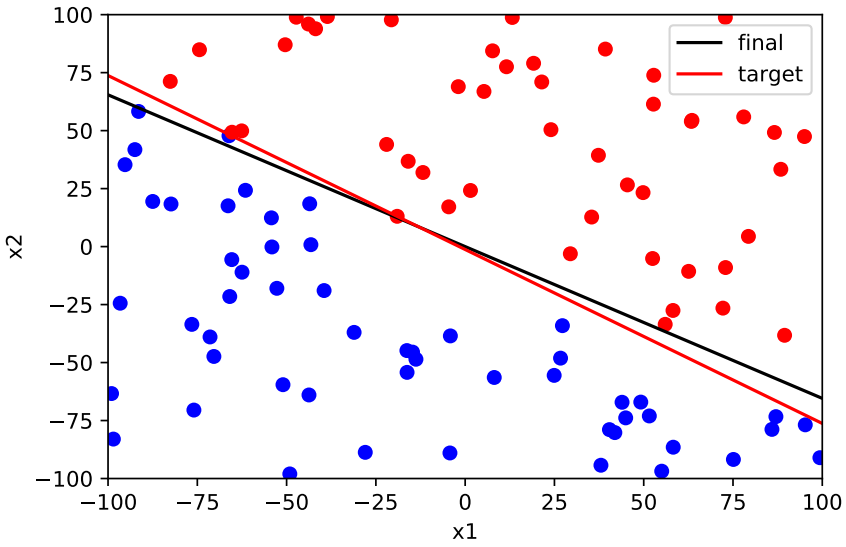
(h)

The more samples I include in the algorithm, the higher accuracy I get, meaning that the final hypothesis is closer to the target function. However, the more samples I include in the algorithm, the larger amount of time I need to implement the algorithm.

The more dimensions of data I include in the algorithm, the larger amount of time I need to implement the algorithm. From (f) I conclude that even a small increase of dimensions of data leads to a significantly larger amount of running time. So, dimensions of data has greater effect on running time than amount of data. The previous questions do not show the relationship of dimension of data and accuracy. I speculate that given the same amount of data, higher dimension may lead to lower accuracy, since higher dimension allows more variability of final hypothesis that correctly classifies data points.

2. Ex 1.5 answer here

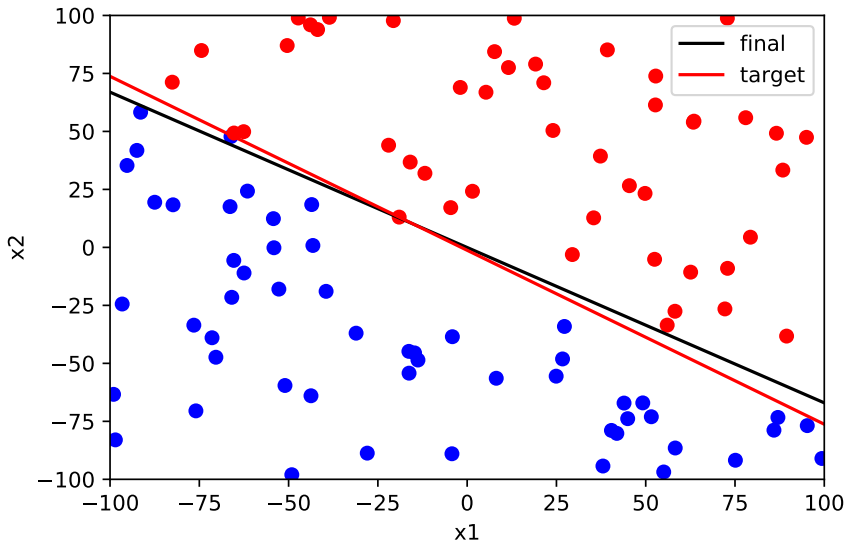
(a)



The error on the test set is 2.36%

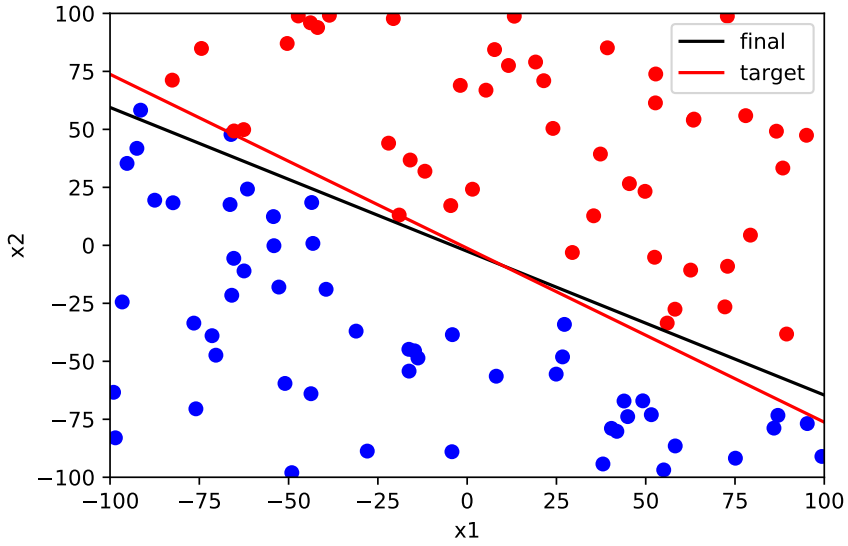
(b)

Problem Set 4



The error on the test set is 1.86%

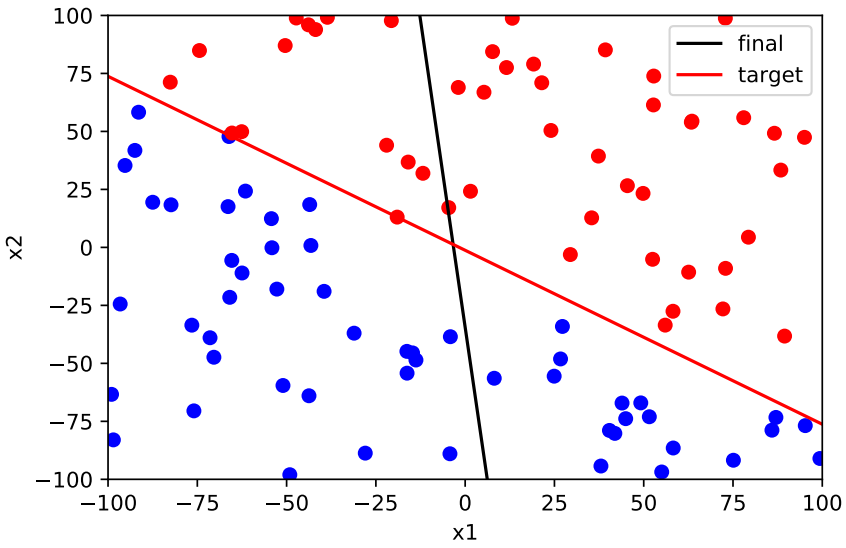
(c)



The error on the test set is 3.26%

(d)

Problem Set 4



The error on the test set is 28.54%

(e) I get the lowest out of sample error rate, when I set the learning rate to 1 and run the algorithm. Actually, when the learning rate is set to 100, 1, or 0.01, the algorithm is performing well (pretty similar). However, when I set learning rate to 0.0001, the out of sample error is big. This is because at each iteration I only adjust weight only a little bit. I speculate if the number of iteration is large enough and I set the learning rate to 0.0001, the algorithm will still perform well.