Homework 03

Philip Hasse

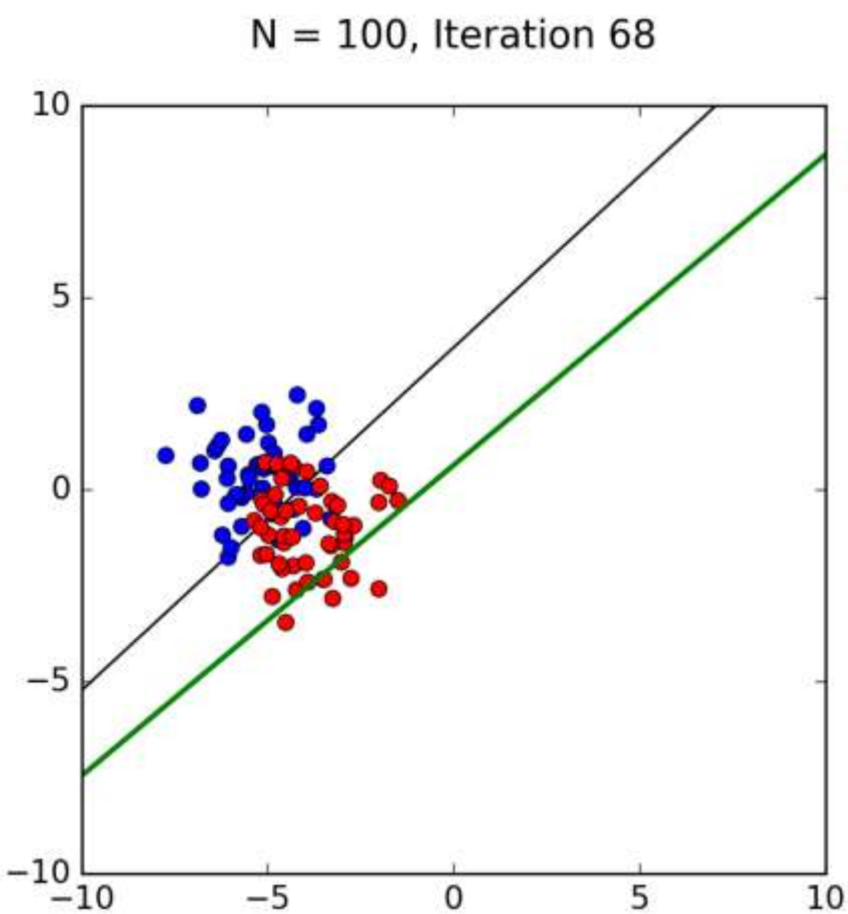
10/11/16

Solution to question 1:

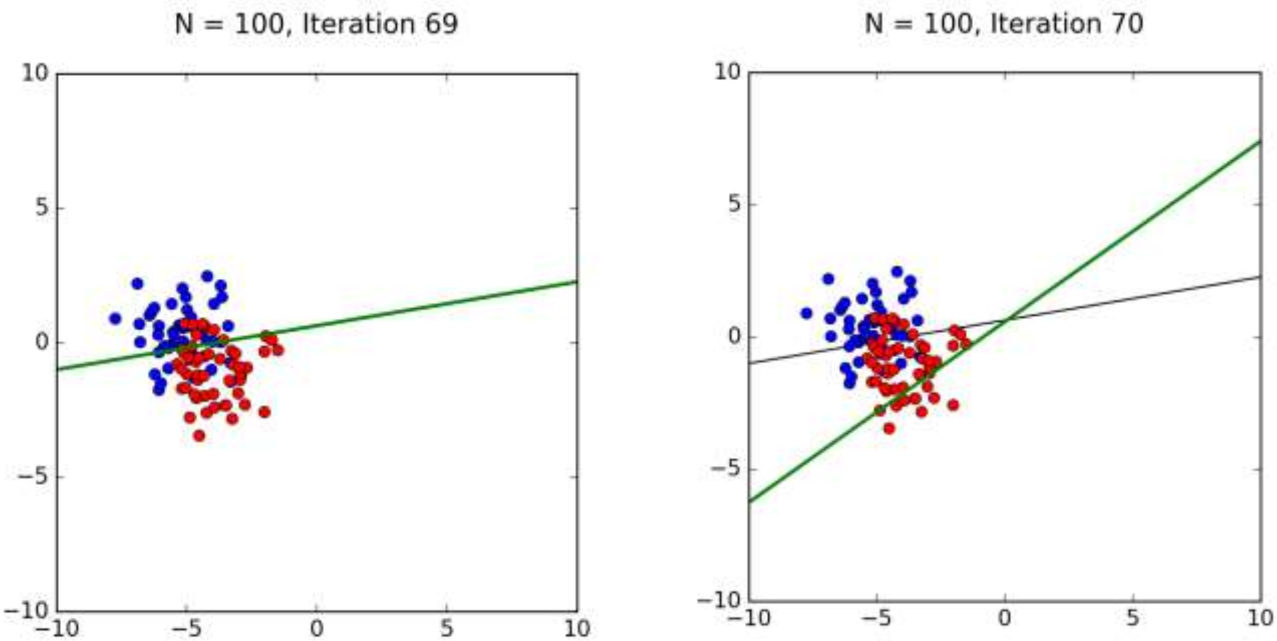
(a) The worst way at finding both the best E_{out} and the lowest amount of computation required is using the Pocket algorithm starting at w=0 because it starts in an arbitrary location and tests each line from there, basically starting at nothing.

(b) The next best way to find both best E_{out} and the lowest amount of computation required is using linear regression. This method is the easiest to perform computation wise, however doesn't always give the best weight needed to separate the data.

(c) The best way out of these three methods to find both best E_{out} and the lowest amount of computation required is using the Pocket algorithm starting for the solution given by linear regression. In the picture below we have the linear regression line still having the best w.



However, in the next iteration the Pocket algorithm finds a line with a better w and replaces the black line with the new best w.



Although this method takes more time to compute the best w, finding a better solution to the problem is more important than giving up a little time.

(outlier) When inserting a significant outlier to the data set the Pocket algorithm will take that point into account and therefore the w will not be as accurate as it should be. Outliers can cause drastic misclassification if not dealt with appropriately.