Daniel Campos October 1st, 2013

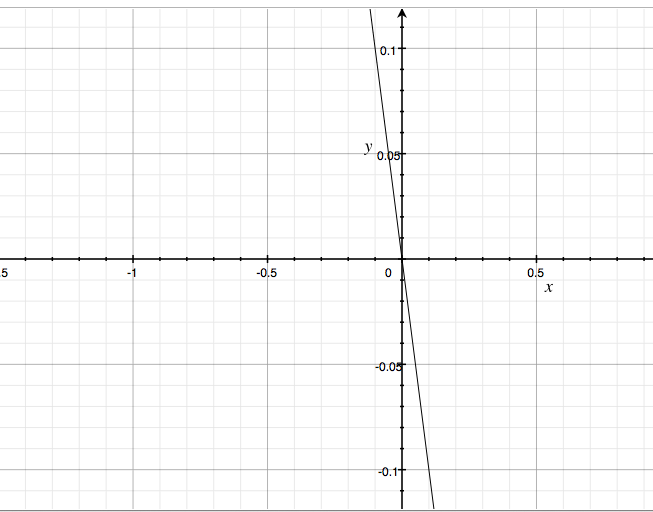
RCS: Campod2 RIN:660996361

Machine Learning CS 4100

Problem Set 5

I worked on thus problem set with Zoe Konrad

1. Exercise 2.8 in LFD(200)
   1. Each  is a final hypothesis output for a data set by a learning algorithm and because of that must belong to the overall hypothesis set H. In that case g(x) is some kind of linear combination of these hypothesis times a certain constant. Since H is closed under a linear combination then g(x) must be one of the infinite amount of H and therefore 
   2. A case where the hypothesis case may not work would be the set as bellow, a simple hypothesis set. . When we run the learning algorithm on both hypothesis sets we will get that our g(x) =0 which is not within the hypothesis set since there is no H(X)=0
   3. Since binary classification is the classification of an item in one of two groups, either has property or not, I believe g(x) will be a binary function since it will classify all points to one side as being part of the function and having the correct value and the others not. For each g(i)(x) there is some kind of line that best classifies the data set. Thus, g(x) is the average of these g(i)(x) and will thus divide all of these functions into two areas and thus be binary.
2. Problem 2.15 in LFD(200)
   1. A monotonic classifier would be as bellow  based in two dimensions where our function is y=-x



* 1. The Mh(N)= , because there is no breakpoint. Assume we have n points. We take a point in this set that each point in monotonically increasing in one component but decreasing in another. We then choose a random point I. Our changes to our data set imply that all components  must be  and . Since none of our data points satisfy this property because they are bigger in one respect and smaller in another, never fully bigger than the previous one. Thus these points have no class for the value of point.

And by extension the VC bound would be

1. Problem 2.24 in LFD(300)
   1. 🡺 and

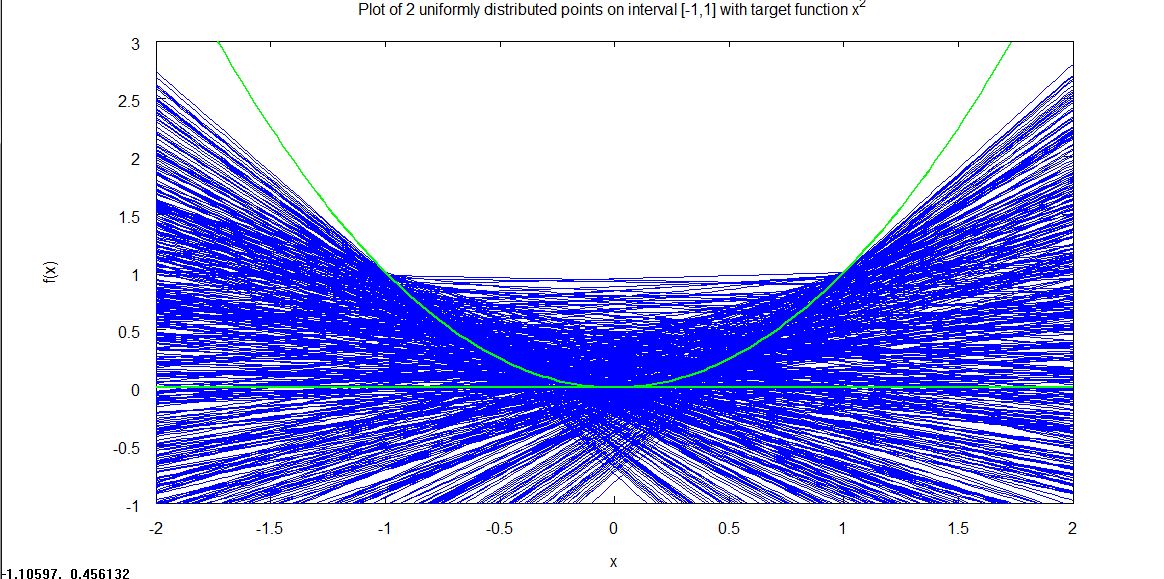
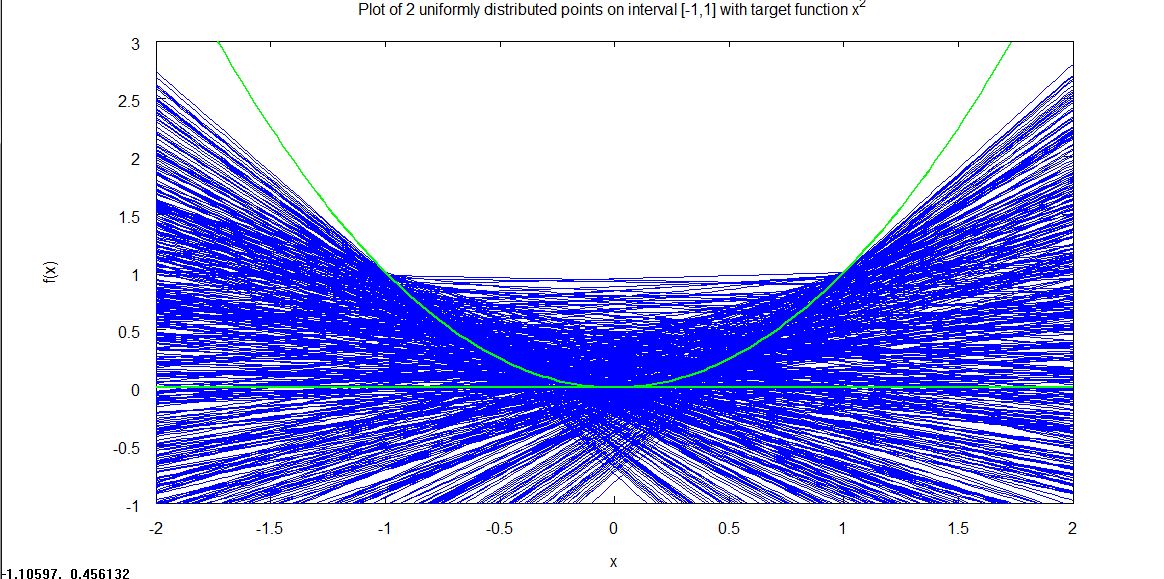


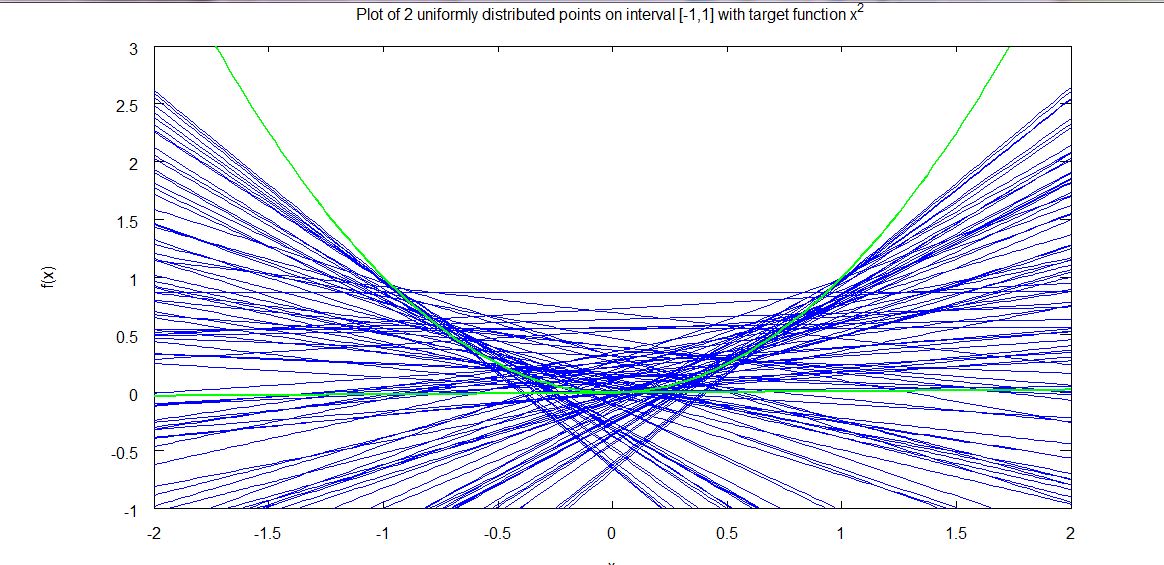
 and  and 

  and 🡪  and our range is -1,1 for both. By this we integrate.

And g(x)=0

* 1. The experiment that I would run to determine g(x), Eout, bias, and Var would be to generate a plot with a line where  from -1 to 1 and plot it and have the program randomly create two points that are on said line. Then, the program would generate the best g(x) for the data. This will be repeated 500 times and add up all the values we calculate the g(x), the Eout, bias and var.
  2. I made my program in octave and plotted the results in my graph. Each blue line represents a g(x), the green line represents f(x) and g(x). I have also listed the var, bias, and eout bellow. I made for 2000 trials, 500 and 100

2000 Trials500 Trials

100 Trials

for 2000 points

* 1. Eout=

Bias= , since my g(x) is basically zero bias is  when we integrate from point -1 to 1 we get that out bias is 1/5

Var=  but once again since my g(x) is zero . We then add out equation in and integrate from -1 to 1 achieving our variance to be 1/3

Eout=bias+variance =1/5+1/3= 8/15