Daniel Campos 660996361

September 12th, 2013 Machine Learning 4000

Problem Set 2

1. Exercise 1.8

If our  then out of 10 marbles we can have either 1 red or no red marbles. Based on the binominal distribution

 where n is amount of red





so the answer is 1\*10-9

1. Exercise 1.9

Since 

We plug in our numbers 🡪

 if we assume that epsilon is .8 then the probability that this is relatively close and shows how small the probability ies

1. Exercise 1.10
   1. The V:First coin was 0.5, random was 0.8 lowest was 0 the 

   4. Both first coin and random coin both obey the bound because they are fixed and do not depend on the data. The minimum coin is different because you are making the distinction of, for each step of 100000, which one is the worst behaved when compared to the average. This in turn will force the lower bounds to be improperly spread.
   5. If a hypothesis h is fixed before generating the data, it obeys the Hoeffding bound. Since we use data to then choose our bin(hypothesis) we cannot use the Hoeffding bound.
2. Exercise 1.11
   1. No we cannot. S cannot guarantee anything about f outside of the dataset
   2. Yes it is possible. Since this set of Data may not actually represent everything else our machine Smart may tell us that all values are +1 when the majority in reality is -1, we just got a set with a high number of +1. The Crazy machine in this case would tell us that all are -1 and thus be more correct outside the Data D.
   3. Given that p=.9





Based in this assumption for p=.9 the Smart learning algorithm will always be better than the crazy one as the smart one will choose h1 and crazy h2

* 1. Any probability of less than .5 will produce better results for the crazy machine. This is true since it starts going for higher probability -1 and works.

1. Exercise 1.12

The best you can promise her would be B since if the data set is meaning ful we can prove her with a final hypothesis and this hypothesis is the most likely one to reach what the real function is out of her data sample. We cannot guarantee it because the sample may not have value, which are equal to the distribution in all data. C is incorrect because we will always produce a hypothesis G, the level of which it is correct will vary but a hypothesis g will always be made.

1. Problem 1.3
   1. Based on w\* being an optimal set of weights, then the outcome y(n) and the predicted outcome w\*x(n) will be the same therefore the product will always be positive sine it effectively becomes  making it always positive.
   2. Base Step

At t=1



w(0)=0 and  =0 and a random misclassified point at time t=0 is and then is correctly classified by w\* at t=1 then  becomes  and the inequality is now . Since we previously define p as the minimum point where it is correctly classified by w\* then 

At time t+1 our inequality becomes

 substitution we get

 this inequality holds true for time t+1 because  is at least p and  is at least p greater than  therefore we conclude that 

* 1. Based on the Euclidean norm  we rewrite our inequality  w we simplify we then get



Based on dot product the inequality becomes



as the dataset is misclassified the expected outcomes will not mat and the signs will not match resulting in the term being negative, thus making the inequality true.

* 1. Based on C we substitute R for  producing  since w(0) =0 and so is its square so  for base case since we showed that  therefore induction proves the question true
  2. In part B we proved that  by taking the quare rout of both terms in d  we divide one by another and get

 and by  the original one is guaranteed to be true

rewriting the equation we simpliofy and solve for t and can conclude that the PLA in fact does converge in a finite amount of time



1. Problem 1.7
   1. 🡪 
      1.  all simulated in octave
         1. 1 Coin 0.59873693923
         2. 1,000 coins
         3. 1,000,000 coins



* + 1. 
       1. 1 Coin 1.024 × 10^-7
       2. 1,000 coins 
       3. 1,000,000 coins 
  1. Graphed bellow very similar to 1.10 part C, almost equal