Problem 1 LFD Problem 1.3

1. By definition is the optimal set of weights such that it separates the data; therefore we have

Therefore,

1. We have the update rule for perceptron learning

By definition we have

We have

Therefore

Because we have , we have

Therefore with this chains of inequalities and by induction we have

1. By the perceptron update rule we have

Because in perceptron learning we have

Because was misclassified we have

Therefore

1. From part (c) we have , which we can further produce

With the chain of inequalities and by induction we have

By definition we have

Therefore we have

1. Using (b) and (d)

First, , we therefore have

Because and in (b) we proved we have

Because in (d) we proved and

Therefore

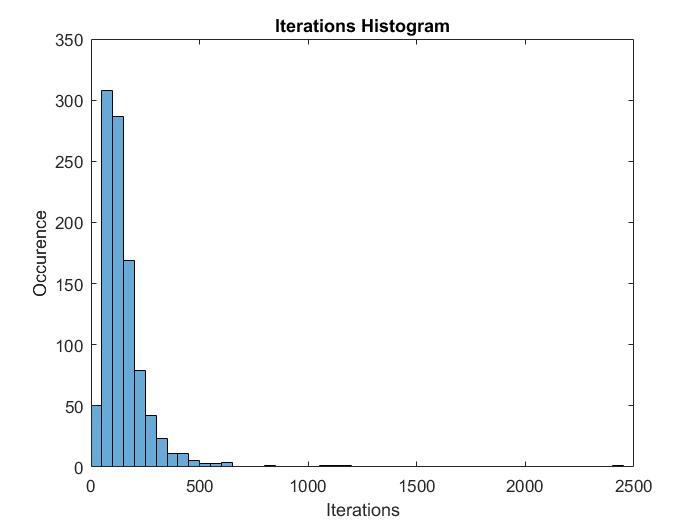
Problem 2

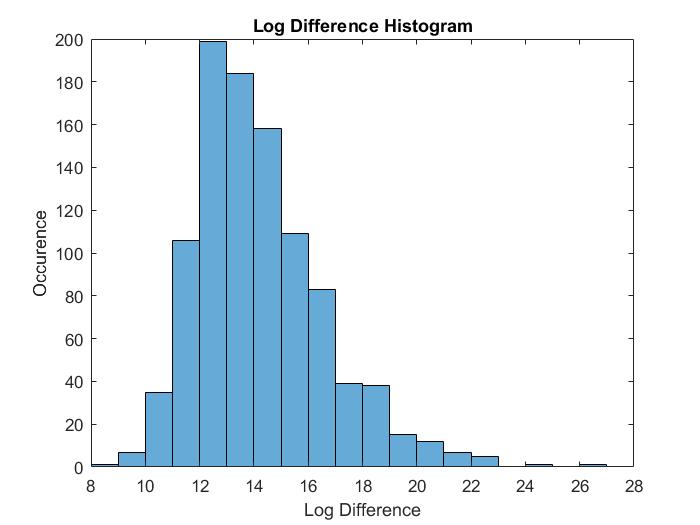
Two graphs are attached below. The first one is the iterations histogram and the second one is the histogram of the log difference between the bound and the actual number of iterations.

Interpretation:

The log-difference has the same distribution of the iteration histogram. They are both normal distributed because the training data and the weights vector are all generated by uniformly distributed random variables.

Another observation is that, the actual number of iteration is always smaller than the bound. And In fact, the bound is very loose as the majority of log difference is ~13, which is about and therefore illustrates that the actual number of iteration is way less than the bound.





Problem 3 LFD Problem 1.7

1. Since , we have the probability that at least one coin will have is equal to

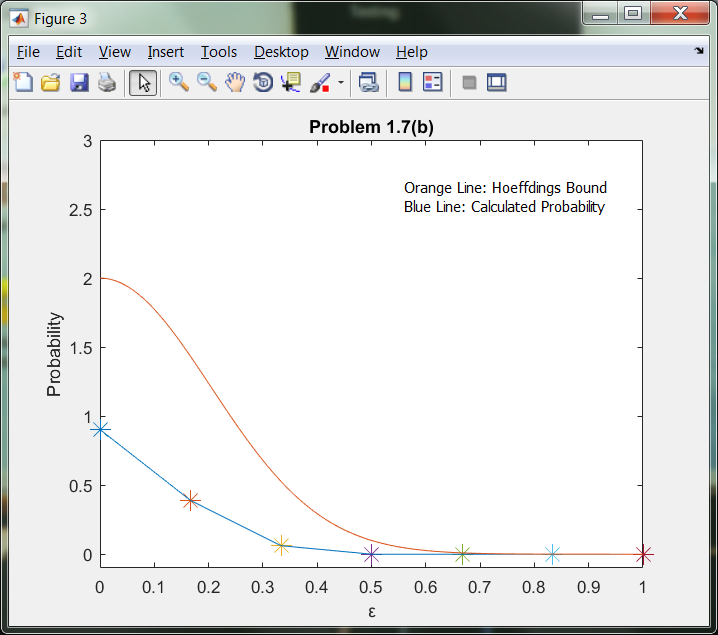
Therefore we have the probability for

And for

1. Hoeffding’s bound:

Calculated probability:

We have the graph generated as shown below. It is very clear that the actual probability is always smaller than the Hoeffding’s bound.



Problem 4 LFD Problem 1.8

1. Let such that all , we then have

Let and , then and

We then have and

Therefore, we have

By definition and is a non-negative random variable, we have

Therefore

1. Let

Plug into the inequality we proved in (a) we have

By the definition, , we have

1. Because are all independent and identically distributed random variables with same mean and same variance

Because , we have and

Plug into the inequality we proved in (b) we have