Problem 1 LFD Problem 1.3

(a) By definition $w^*$ is the optimal set of weights such that it separates the data; therefore we have

$$sign(w^{*T}x_n) = sign(y_n) \text{ for } 1 \leq n \leq N$$

Therefore,

$$p = \min_{1 \leq n \leq N} y_n(w^{*T}x_n) > 0$$

(b) We have the update rule for perceptron learning

$$w(t+1) = w(t) + y(t)x(t) \rightarrow w(t) = w(t-1) + y(t-1)x(t-1)$$

$$w^T(t)w^* = \left(w(t-1) + y(t-1)x(t-1)\right)^T w^* = w^T(t-1)w^* + y(t-1)^T(x(t-1)^T)w^*$$

$$= w^T(t-1)w^* + y(t-1)(w^{*T}x(t-1))$$

By definition $p = \min_{1 \leq n \leq N} y_n(w^{*T}x_n)$ we have $p \leq y_n(w^{*T}x_n)$

We have

$$y(t-1)\left(w^{*T}x(t-1)\right) \geq p$$

Therefore

$$w^T(t)w^* = w^T(t-1)w^* + y(t-1)\left(w^{*T}x(t-1)\right) \geq w^T(t-1)w^* + p$$

Because we have $w^T(t)w^* \geq w^T(t-1)w^* + p$, we have

$$w^T(t-1)w^* \geq w^T(t-2)w^* + p$$

$$w^T(t-2)w^* \geq w^T(t-3)w^* + p$$

$$\dots$$

$$w^T(t-(t-1))w^* \geq w^T(t-t)w^* + p$$

Therefore with this chains of inequalities and by induction we have

$$w^T(t)w^* \geq tp$$

(c) By the perceptron update rule $w(t) = w(t-1) + y(t-1)x(t-1)$ we have

$$\left||w(t)|\right|^2 = \sum w^2(t)_i = \sum(w(t-1)_i + y(t-1)x(t-1)_i)^2$$

$$= \sum(w^2(t-1)_i + y^2(t-1)x^2(t-1)_i + 2w(t-1)_i y(t-1)x(t-1)_i)$$

$$= \sum w^2(t-1)_i + \sum y^2(t-1)x^2(t-1)_i + \sum 2w(t-1)_i y(t-1)x(t-1)_i$$

Because in perceptron learning $y = \pm 1$ we have $y^2 = 1$

$$\left||w(t)|\right|^2 = \sum w^2(t-1)_i + \sum x^2(t-1)_i + 2\sum w(t-1)_i y(t-1)x(t-1)_i$$

$$= \left||w(t-1)|\right|^2 + \left||x(t-1)|\right|^2 + 2\sum w(t-1)_i y(t-1)x(t-1)_i$$

Because $x(t-1)$ was misclassified we have $\sum w(t-1)_i y(t-1)x(t-1)_i \leq 0$

Therefore $||w(t)||^2 \leq ||w(t-1)||^2 + ||x(t-1)||^2$

(d) From part (c) we have $||w(t)||^2 \leq ||w(t-1)||^2 + ||x(t-1)||^2$, which we can further produce

$$||w(t-1)||^2 \leq ||w(t-2)||^2 + ||x(t-2)||^2$$

$$\ldots$$

$$||w(t-(t-1))||^2 \leq ||w(t-t)||^2 + ||x(t-t)||^2$$

With the chain of inequalities and by induction we have

$$||w(t)||^2 \leq \sum_{i=1}^{t-1} ||x(i)||^2 + ||x(t-1)||^2$$

By definition $R = \max_{1\leq n\leq N} ||x_n||$ we have $R^2 \geq ||x(i)||$ for any $i$

Therefore we have

$$||w(t)||^2 \leq (t-1)R^2 + R^2 = tR^2$$

(e) Using (b) and (d)

First, $\sqrt{t} \geq 0, p > 0, R > 0$, we therefore have

$$\frac{w^T(t)}{||w(t)||} w^* \geq \sqrt{t} \times \frac{p}{R} \to \frac{w^T(t)^2}{||w(t)||^2} w^{*2} \geq t \times \frac{p^2}{R^2}$$

$$w^{T2}(t) w^{*2} R^2 \geq tp^2 ||w(t)||^2 \to [w^T(t)w^*]^2 tR^2 \geq (tp)^2 ||w(t)||^2$$

Because $p, t > 0$ and in (b) we proved $w^T(t)w^* \geq tp$ we have $[w^T(t)w^*]^2 \geq (tp)^2$

Because in (d) we proved $tR^2 \geq ||w(t)||^2$ and $||w(t)||^2 \geq 0$

Therefore $[w^T(t)w^*]^2 tR^2 \geq (tp)^2 ||w(t)||^2 \to \frac{w^T(t)}{||w(t)||} w^* \geq \sqrt{t} \times \frac{p}{R}$

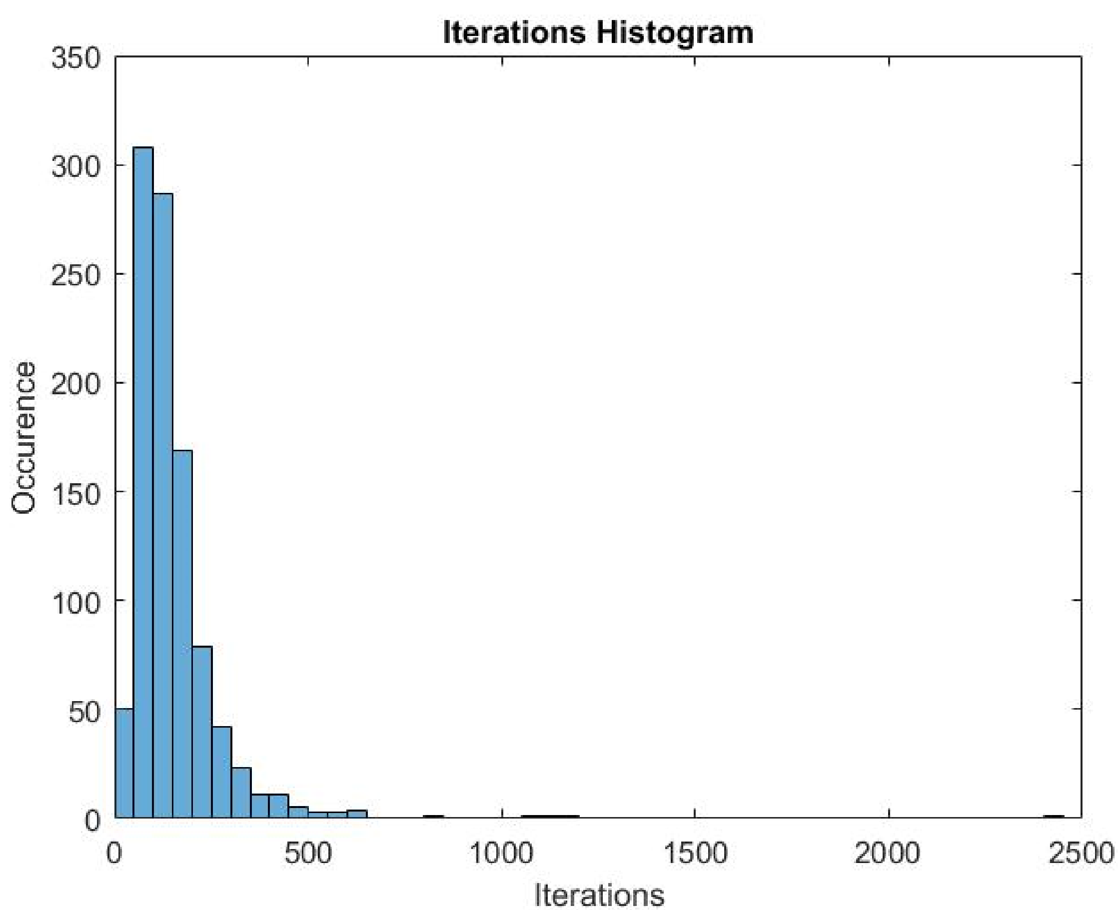
Problem 2

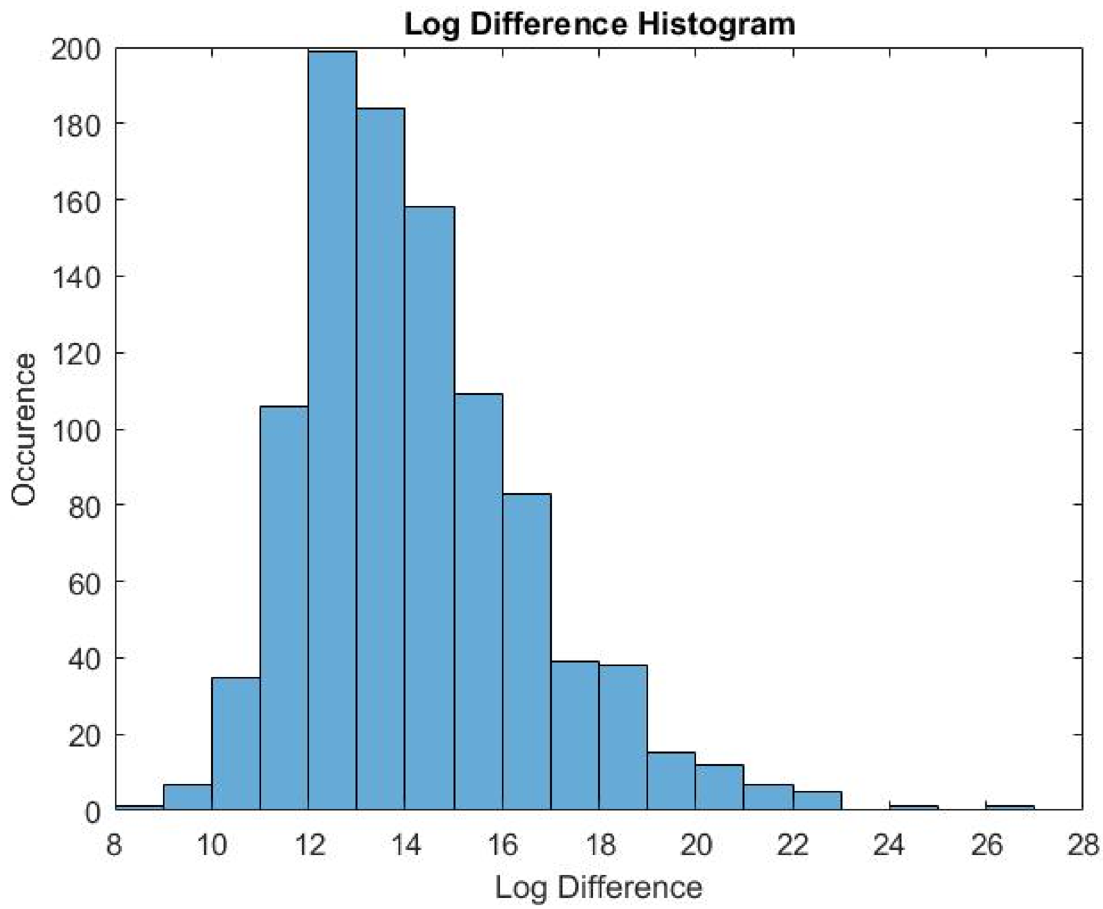
Two graphs are attached below. The first one is the iterations histogram and the second one is the histogram of the log difference between the bound and the actual number of iterations.

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 $e^{13} \approx 162000$ and therefore illustrates that the actual number of iteration is way less than the bound.




Log Difference Histogram
Occurence
Log Difference
0
20
40
60
80
100
120
140
160
180
200
8
10
12
14
16
18
20
22
24
26
28

Problem 3 LFD Problem 1.7

(a) Since $P[k < 1|N, \mu] + P[k \geq 1|N, \mu] = 1$, we have the probability that at least one coin will have $v = 0$ is equal to

$$P[k \geq 1|N, \mu] = 1 - P[k = 0|N, \mu]$$

Therefore we have the probability for $\mu = 0.05$

$P_1 = P[k = 0|10, 0.05] = 0.5987$

$P_{1000} = 1 - P[k \geq 1|10, 0.05]^{1000} = 1 - (1 - 0.5987)^{1000} = 1$

$P_{1000000} = 1 - P[k \geq 1|10, 0.05]^{1000000} = 1 - (1 - 0.5987)^{1000000} = 1$

And for $\mu = 0.8$

$P_1 = P[0|N, 0.8] = \frac{1}{9765625} = 0$

$P_{1000} = 1 - P[k \geq 1|10, 0.8]^{1000} = 1 - (1 - \frac{1}{9765625})^{1000} = 0.0001$

$P_{1000000} = 1 - P[k \geq 1|10, 0.8]^{1000000} = 1 - (1 - \frac{1}{9765625})^{1000000} = 0.0973$

(b) Hoeffding's bound: $y = 2 * e^{-2 \times 6 \times \varepsilon^2}$

Calculated probability: $y = 1 - \int_{k=3-\varepsilon}^{3+\varepsilon} \binom{6}{k} \mu^k (1 - \mu)^{6-k}$

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

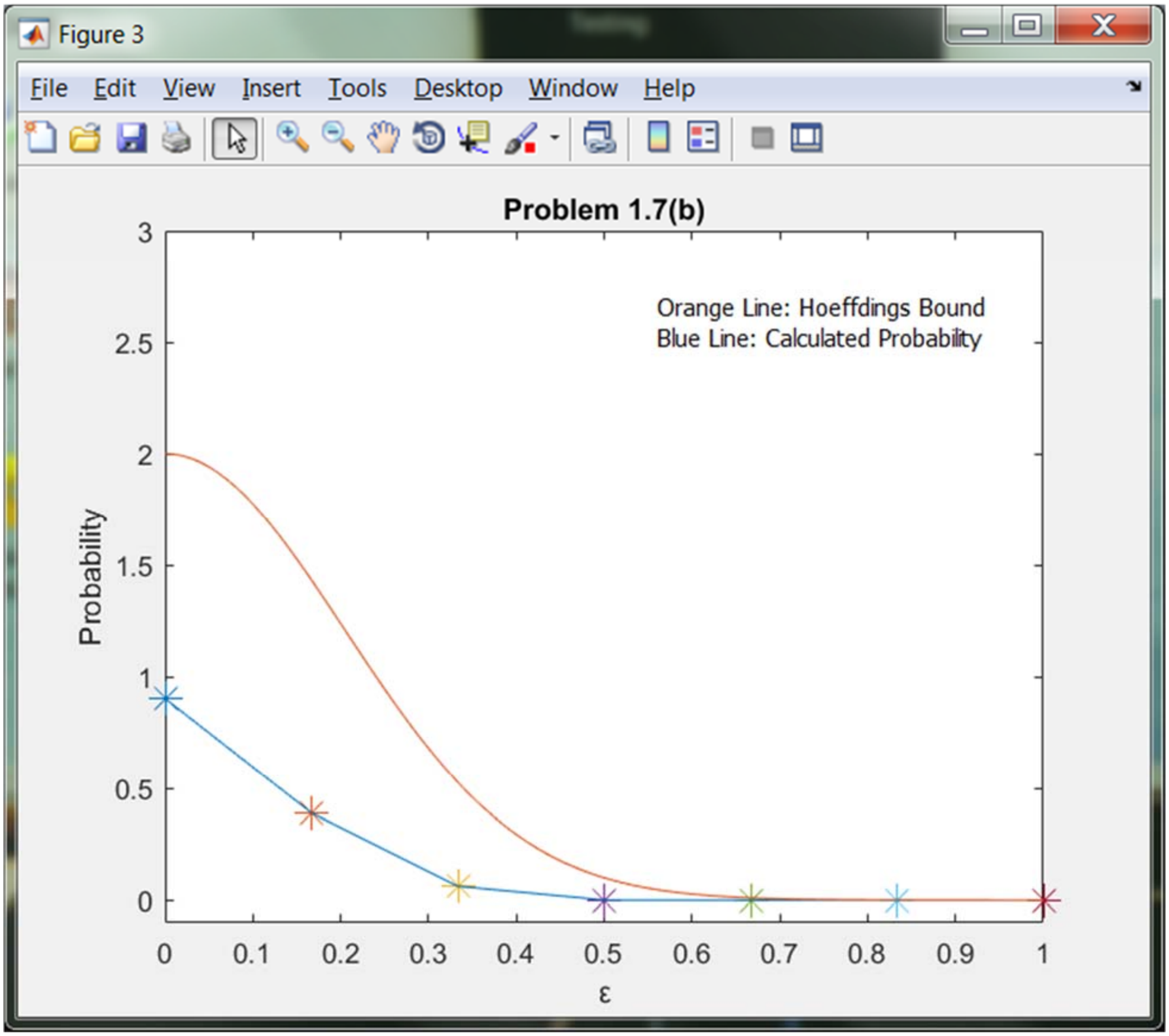


Figure 3
File Edit View Insert Tools Desktop Window Help
Problem 1.7(b)
Orange Line: Hoeffdings Bound
Blue Line: Calculated Probability
Probability
ε
0
0.5
1
1.5
2
2.5
3
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Problem 4 LFD Problem 1.8

(a) Let $t^* \subseteq t$ such that all $t^* = \{t_i | t_i \geq \alpha\}$, we then have $t - t^* = \{t_i | t_i < \alpha\}$

Let $|t| = m$ and $|t^*| = n$, then $m \geq n$ and $|t - t^*| = m - n$

We then have $P[t \geq \alpha] = \frac{n}{m}$ and $E[t] = \frac{n}{m} E[t^*] + \frac{m-n}{m} E[t - t^*]$

Therefore, we have

$$\frac{E[t]}{\alpha} = \frac{n}{m}\frac{E[t^*]}{\alpha} + \frac{m-n}{m}\frac{E[t-t^*]}{\alpha}$$

By definition $t^* = \{t_i | t_i \geq \alpha\}$ and $t$ is a non-negative random variable, we have

$$\frac{E[t^*]}{\alpha} \geq 1$$

$$\frac{E[t-t^*]}{\alpha} \geq 0$$

Therefore

$$\frac{E[t]}{\alpha} = \frac{n}{m}\frac{E[t^*]}{\alpha} + \frac{m-n}{m}\frac{E[t-t^*]}{\alpha} \geq \frac{n}{m} = P[t \geq \alpha]$$

(b) Let $t = (u - \mu)^2$

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

$$P[t \geq \alpha] \leq \frac{E[t]}{\alpha} \rightarrow P[(u-\mu)^2 \geq \alpha] \leq \frac{E[(u-\mu)^2]}{\alpha}$$

By the definition, $\sigma^2 = variance = E[(u-\mu)^2]$, we have

$$P[(u-\mu)^2 \geq \alpha] \leq \frac{E[(u-\mu)^2]}{\alpha} = \frac{\sigma^2}{\alpha}$$

(c) Because $u_1, u_2, \ldots, u_n$ are all independent and identically distributed random variables with same mean $\mu$ and same variance $\sigma^2$

Because $u = \frac{1}{N}\sum_{n=1}^{N} u_n$, we have $\mu_u = \frac{1}{N}\sum_{n=1}^{N} \mu = \mu$ and $\sigma_u^2 = \sum_{n=1}^{N} \frac{1}{N^2}\sigma^2 = \frac{1}{N}\sigma^2$

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

$$P[(u-\mu_u)^2 \geq \alpha] = \frac{{\sigma_u}^2}{\alpha} \rightarrow P[(u-\mu)^2 \geq \alpha] = \frac{\sigma^2}{N\alpha}$$