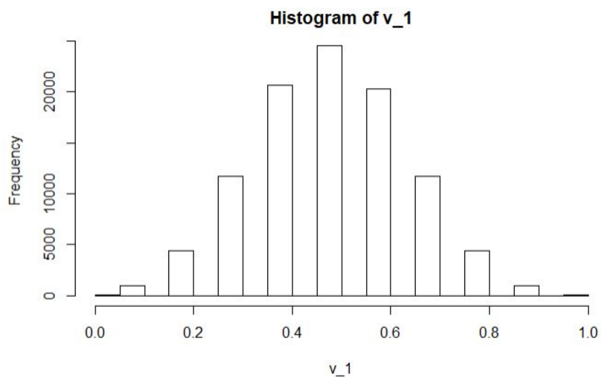
Homework 2

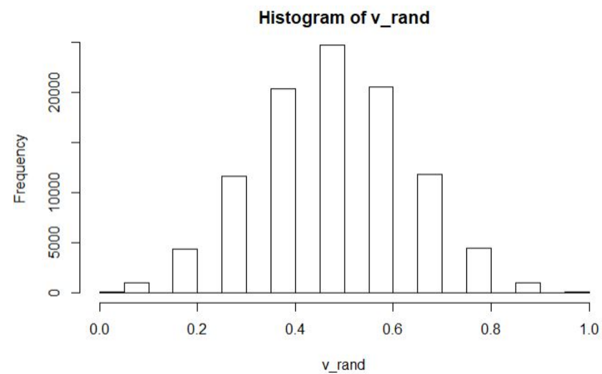
Karan Sarkar
sarkak2@rpi.edu

September 17, 2019

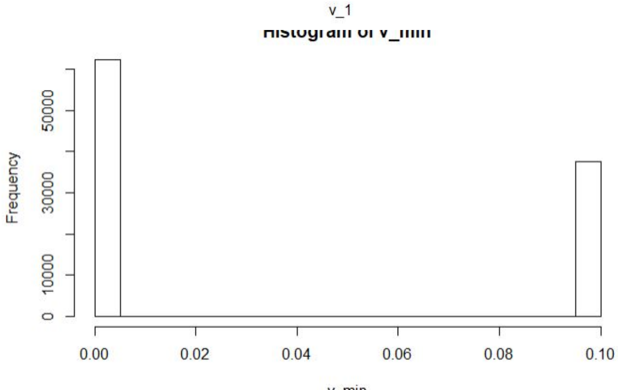
Exercise 1.8. There are two possibilites: 0 red or 1 red. 0 red happens with probability $(0.1)^{10}$. 1 red happens wih probability $10 \cdot (0.1)^9 \cdot 0.9$. THe overall probability is $9.1 \cdot 10^{-9}$.
Exercise 1.9. Using the Hoeffding bound, we have $P[|\nu - \mu| > 0.8] \leq 2e^{-2(0.8)^2 \cdot 10} \approx 5.521 \cdot 10^{-6}$. This is still a small probability but larger than the value computed in Exercise 1.8.
Exercise 1.10.

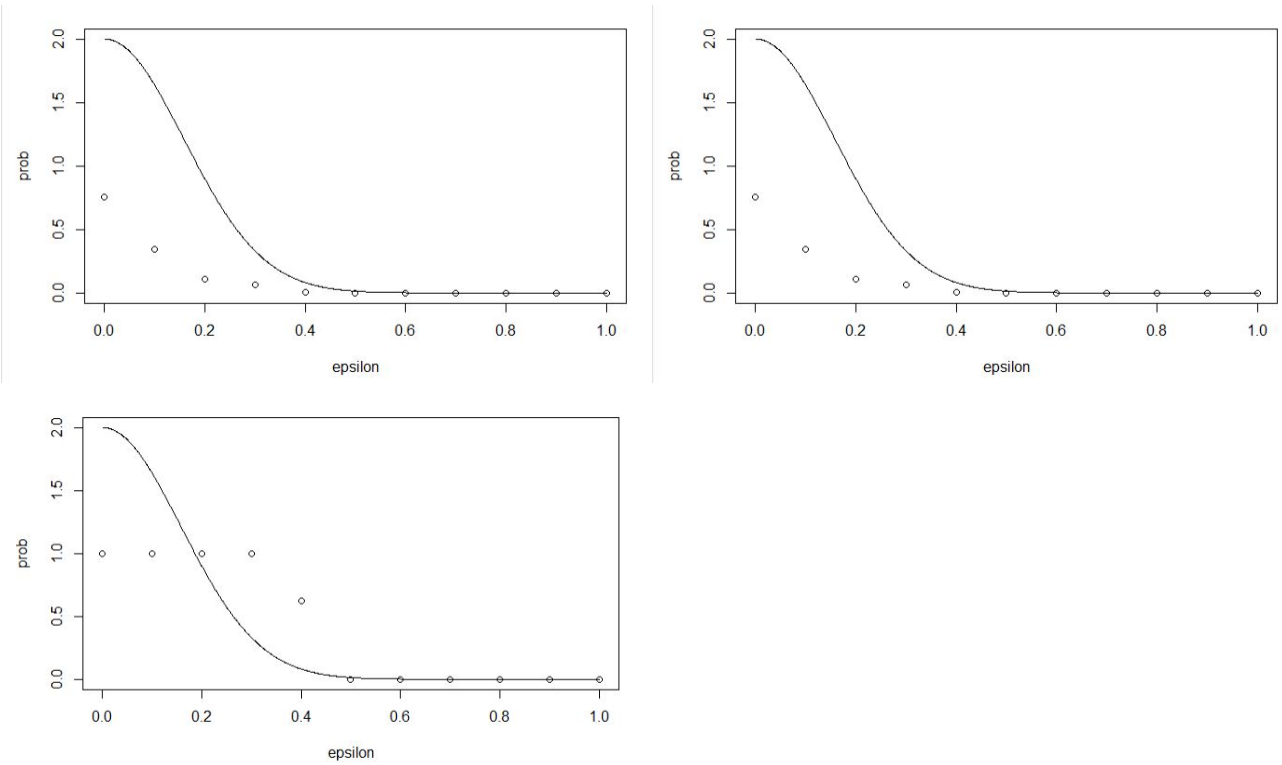
(a) The value of μ is 0.5 for all coins.





(b)



(c) 

The first two are for ν_1 and ν_{rand}. The last is for ν_{min}.

(d) We can see that that the Hoeffding bound holds for ν_1 and ν_{rand}. It does not hold for ν_{min}.

(e) Our 1000 different coins are the multiple bins. We cannot use the Hoeffding bound on a non-fixed bin like $\nu_m in$ which performed the best.

Exercise 1.11.

(a) No S cannot guarantee to produce a better hypothesis.

(b) Yes, It is possible C produces a better hypothesis. It is possible we were just unlucky.

(c) We are looking for the probability that S will choose a better hypothesis than C. This will happen if a minority of the points are positive. This probability is equal to

$$\sum_{i=0}^{12} \binom{25}{i} (0.9)^i (0.1)^{25-i} \approx 1.62 \cdot 10^{-7}$$

.

(d) There is no value of μ that C is more likely to produce a better hypothesis.

Exercise 1.12. We can only guarantee (c).
Problem 1.3.

(a) We know that $\mathbf{x}$ is properly classified by $\mathbf{w}^*$ because the data is linearly separable. This yields two cases. First, suppose that we have a true positive. This means that $h(\mathbf{x}) = \text{sign}(\mathbf{w}^{*T}\mathbf{x}) > 0$ while $y > 0$. Thus, $\mathbf{w}^{*T}\mathbf{x} > 0$. Thus, $y\mathbf{w}^{*T}\mathbf{x} > 0$. Second, suppose that we have a true negative. This means that $h(\mathbf{x}) = \text{sign}(\mathbf{w}^{*T}\mathbf{x}) < 0$ while $y < 0$. Thus, $\mathbf{w}^{*T}\mathbf{x} < 0$. Thus, $y\mathbf{w}^{*T}\mathbf{x} > 0$. The result thus holds in both cases. Because this holds for all $\mathbf{x}$ it holds for the minimum. Thus, $\rho = \min_{1 \le n \le N} y_n \mathbf{w}^{*T}\mathbf{x_n} > 0$

(b) We will prove $\mathbf{w}^T(t)\mathbf{w}^* \geq \mathbf{w}^T(t-1)\mathbf{w}^* + \rho$. Consider the quantity $\mathbf{w}^T(t)\mathbf{w}^*$. From the definition of the update rule we have:

$$\begin{aligned}
\mathbf{w}^T(t)\mathbf{w}^* &= [\mathbf{w}(t-1) + y(t-1)\mathbf{x}(t-1)]^T\mathbf{w}^* \\
&= \mathbf{w}^T(t-1)\mathbf{w}^* + y(t-1)\mathbf{w}^{*T}\mathbf{x}(t-1) \\
&\geq \mathbf{w}^T(t-1)\mathbf{w}^* + \rho
\end{aligned}$$

We will now show by induction that $\mathbf{w}^T(t)\mathbf{w}^* \geq t\rho$. For the base case, we will show that $\mathbf{w}^T(0)\mathbf{w}^* \geq 0$. This is true because $\mathbf{w}(0) = 0$. For the inductive step, we will assume $\mathbf{w}^T(k)\mathbf{w}^* \geq k\rho$. We will try to prove $\mathbf{w}^T(k+1)\mathbf{w}^* \geq (k+1)\rho$. This is true as $\mathbf{w}^T(k+1)\mathbf{w}^* \geq \mathbf{w}^T(k)\mathbf{w}^* + \rho \geq k\rho + \rho = (k+1)\rho$. Thus, we have that for all t, $\mathbf{w}^T(t)\mathbf{w}^* \geq t\rho$.

(c) Consider the quantity $||\mathbf{w}(t)||^2$. By the update step, we have:

$$\begin{aligned}
||\mathbf{w}(t)||^2 &= ||\mathbf{w}(t-1) + y(t-1)\mathbf{x}(t-1)||^2 \\
&= ||\mathbf{w}(t-1)||^2 + 2y(t-1)\mathbf{w}(t-1)^T\mathbf{x}(t-1) + y(t-1)^2||\mathbf{x}(t-1)||^2 \\
&\leq ||\mathbf{w}(t-1)||^2 + ||\mathbf{x}(t-1)||^2
\end{aligned}$$

Note that $y(t-1)\mathbf{w}(t-1)^T\mathbf{x}(t-1) < 0$ when a point is misclassified as follows from (a). Thus, it follows that $||\mathbf{w}(t)||^2 \leq ||\mathbf{w}(t-1)||^2 + ||\mathbf{x}(t-1)||^2$.

(d) We will now show by induction that $||\mathbf{w}(t)||^2 \geq tR^2$. For the base case, we will show that $||\mathbf{w}(0)||^2 \geq 0$. This is true because $\mathbf{w}(0) = 0$. For the inductive step, we will assume $||\mathbf{w}(k)||^2 \geq kR^2$. We will try to prove $||\mathbf{w}(k+1)||^2 \geq (k+1)R^2$. This is true as $||\mathbf{w}(k+1)||^2 \geq ||\mathbf{w}(k)||^2 + ||\mathbf{x}(k)||^2 \geq kR^2 + R^2 = (k+1)R^2$. Thus, we have that for all t, $||\mathbf{w}(t)||^2 \geq tR^2$.

(e) Note that from (d), we have $||\mathbf{w}(t)|| \geq \sqrt{t}R$. Now consider the quantity $\frac{\mathbf{w}^T(t)\mathbf{w}^*}{||\mathbf{w}(t)||}$.

$$\frac{\mathbf{w}^T(t)\mathbf{w}^*}{||\mathbf{w}(t)||} \geq \frac{t\rho}{\sqrt{t}R} = \sqrt{t}\frac{\rho}{R}$$

Consider the angle between $\mathbf{w}(t)$ and $\mathbf{w}^*$. The cosine of this angle is less than 1. Thus, we have:

$$\begin{aligned}
\frac{\mathbf{w}^T(t)\mathbf{w}^*}{||\mathbf{w}(t)||||\mathbf{w}^*||} &\leq 1 \\
\sqrt{t}\frac{\rho}{R||\mathbf{w}^*||} &\leq 1 \\
t &\leq \frac{R^2||\mathbf{w}^*||^2}{\rho^2}
\end{aligned}$$

Problem 1.7.

(a) When $\mu = 0.05$, for 1 coin, 1000 coins and 1000000 coins, we have probabilities of 0.60, 1 and 1 respectively. When $\mu = 0.8$, for 1 coin, 1000 coins and 1000000 coins, we have probabilities of $1.024 \cdot 10^{-7}$, $1.024 \cdot 10^{-4}$ and 0.097 respectively.

(b)