Machine Learning From Data HW2

Shane O'Brien

September 2017

Exercise 1.8

To calculate the $P[\nu \leq 0.1]$, we add the probability $P[\nu = 0.1]$ and $P[\nu = 0.0]$

$P[\nu = 0.1] = (0.9 * 0.1^9) * \binom{10}{1} = 9 * 10^{-9}$

$P[\nu = 0.0] = (0.1^{10}) * \binom{10}{0} = 1 * 10^{-10}$

$P[\nu \leq 0.1] = P[\nu = 0.1] + P[\nu = 0.0] = 9.1 * 10^{-9}$

Exercise 1.9

$P[|\nu - \mu| > \mathcal{E}] \leq 2 * e^{-2*N*(\mathcal{E})^2)}$

$\leq 2 * e^{-2*100*(0.8)^2)}$

$\leq 5.52 * 10^{-6}$

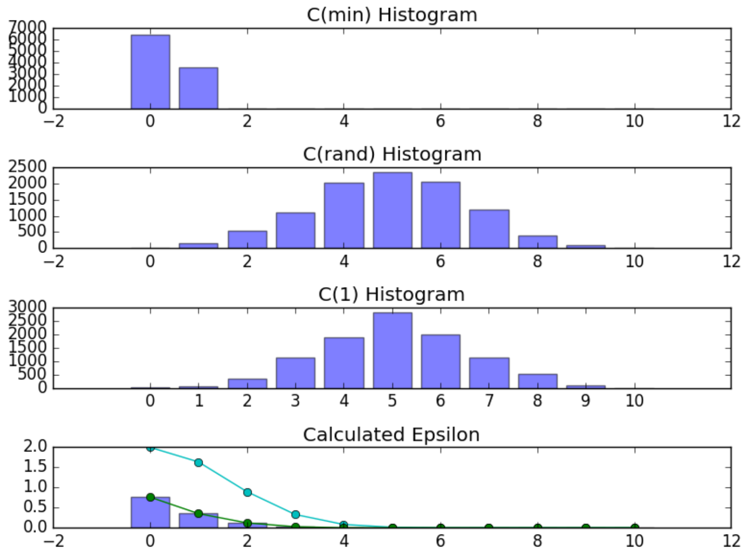
Exercise 1.10

a

The μ for all three coins is 0.5

b

The image of this test is below. I ran just 10,000 tests because I am writing code on a very weak machine. 100,000 tests would take well over a day.



The top graph is a histogram of C_{min}. The second from the top is a histogram of C_{rand}. The third from the top is a histogram of C_1. And the bottom one is important in part c.

c

Look at the figure above to see the calculated epsilon. As expected, the plot for the estimated P$[|\nu - \mu| > \mathcal{E}]$ is below the Hoeffding bound.

d

The coins that were selected as the minimum do not obey the Hoeffding bound. To be applicable to the Hoeffding bound, the selection must be completely random. Obviously, the minimum selections were not random.

e

The bins in Figure 1.10 are valid, unlike the minimum selection from this problem. Although the results could be similar, how they are achieved is just as important. In Figure 1.10, the marbles are randomly selected. In the minimum selection, there is a clear pattern to selection.

Exercise 1.11

a

No, nothing is guaranteed outside $\mathcal{D}$

b

Yes, it is possible. We can always get 'fooled' by the data set. Nothing is guaranteed.

c
I am going to provide the exact probabilistic answer, and then the Hoeffding bound

If p = 0.9, P[S will produce better hypothesis than C] =

$= \Sigma_{n=13}^{25} * \binom{25}{n} * (0.9)^n * (0.1)^{25-n}$

$= 0.9999998$

Now, using the Hoeffding bound:

$P[|\nu - \mu| > \mathcal{E}] \leq 2 * e^{-2*N*(\mathcal{E})^2)}$

$\leq 2 * e^{-2*25*(0.4)^2)}$

≤ 0.00671

So the Hoefding bound states...
P[S will produce better hypothesis than C] ≥ 0.999329

Exercise 1.12

Our best option is **c**. As always, there is never a guarantee. This means that we can't even guarantee that we will find a g.

Problem 1.3

a
By definition, any data point x_n in $y_n(w^{*T} x_n)$ is ¿ 0. This is because w^* is completely correct

b
Known before:
$w^T(t-1) = w^T(t) - y(t)x^T(t)$

and from part a:

$w^T(t) - y(t)x^T(t)w^* + \rho \leq w^T(t)w^*$

Base Case:
$t = 0$
$w^T(0)w^* \geq 0p \longleftrightarrow 0 \geq 0$

Induction:
We assume: $w^T(t+1)w^* \geq (t+1)\rho$

So then it follows that: $(t+1)\rho = t\rho + \rho \leq w^T(t+1)w^*$

c

Consider the case: $y(t+1) = +1$, where
$\|w(t)\|^2 = \|w(t-1)\|^2 + \|x(t+1)\|^2 - 2 * \|w(t-1)\| * \|x(t-1)\| cos(\theta)$
Therefore:
$cos(\theta) \geq 0 \longleftrightarrow -2\|w(t-1)\| * \|x(t-1)\| cos(\theta) \leq 0.$

This case: $y(t-1) = -1$, is the same:

$\|w(t)\|^2 = \|w(t-1)\|^2 + \|x(t+1)\|^2 - 2 * \|w(t-1)\| * \|x(t-1)\| cos(\theta) \leq \|w(t-1)\|^2 + \|x(t+1)\|^2$

d

Base Case:
t = 0, so therefore:
$\|w(0)\|^2 \leq 0 * \mathcal{R}^2 \longleftrightarrow 0 \leq 0$

Induction:
Assume: $\|w(t+1)\|^2 \leq (t+1) * \mathcal{R}^2$

So therefore, with help from part c:

$(t+1) * \mathcal{R}^2 = t * \mathcal{R}^2 + \mathcal{R}^2 \geq \|w(t)\|^2 + \mathcal{R}^2 \geq \|w(t)\|^2 + \|x(t-1)\|^2 \geq \|w(t+1)\|^2$

e

We use our proofs from part b and part d.

$\frac{w^T(t)}{\|w(t)\|} w^* \geq \frac{w^T(t-1)w^* + \rho}{\|w(t)\|} \geq \frac{t\rho}{\|w(t)\|}$

$\frac{t\rho}{|(t)\|} \geq \frac{t\rho}{\sqrt{t}\mathcal{R}} = \frac{\sqrt{t}\rho}{\mathcal{R}}$

So we have: $\frac{w^T(t)}{\|w(t)\|} w^* \geq \sqrt{t} * \frac{\rho}{\mathcal{R}}$

Consider:

$\frac{w^T(t)}{\|w(t)\|} w^* = \|w^*\| cos(\theta)$

In this case θ is the specific angle between $w^T(t)$ and w^*

Based on this, we have:
$\sqrt{t} * \frac{\rho}{\mathcal{R}} \geq 0 \frac{w^T(t)}{\|w(t)\|} w^* \geq 0 \longleftrightarrow \|w^*\| cos(\theta) \geq 0 \longleftrightarrow cos(\theta) 0 \longleftrightarrow 0 \leq 90$

If $\|w^*\| cos(\theta)$ increase, so does $cos(\theta)$. This means θ goes down. This leads to $w(t)$ getting infinitely close to w^*

The final bound for t is:

$t \leq \frac{\mathcal{R}^2 \|w^*\|^2}{\rho^2}$

Problem 1.7

a

$\mu = 0.05$, coin = 1
P[one coin has no heads] = $(0.95)^{10} = 0.5987$

$\mu = 0.05$, coin = 1,000
P[getting at least one head] = 0.401263
P[at least 1 coin of 1000 with 0 heads] = 1 - $(0.401263)^{1000} = 1 - 2.68647 * 10^{-397}$

$\mu = 0.05$, coin = 1,000,000
P[getting at least one head] = 0.401263
P[at least 1 coin of 1000000 with 0 heads] = 1 - $(0.401263)^{1000000} = 1 - (0.401263)^{1000000}$

$\mu = 0.80$, coin = 1
P[one coin has no heads] = $(0.2)^{10} = 1.024 * 10^{-7}$

$\mu = 0.80$, coin = 1,000
P[getting at least one head] = $1 - 1.024 * 10^{-7}$
P[at least 1 coin of 1000 with 0 heads] = $1 - (1 - 1.024 * 10^{-7})^{1000}] = 1 - 0.999898 = 0.00102$

' **$\mu = 0.80$, coin = 1,000,000**
P[getting at least one head] = $1 - 1.024 * 10^{-7}$
P[at least 1 coin of 1000000 with 0 heads] = $1 - (1 - 1.024 * 10^{-7})^{1000000}] = 1 - 0.902668 = 0.097332$

b

First, we look at P[$|\nu - \mu| > \mathcal{E}$] for just one coin.

P[$|\nu_i - \mu_i| > 0.0$]] = P[0,1,2,4,5,6 heads] = $\frac{44}{64}$

$\mathrm{P}[|\nu_i - \mu_i| > 0.1]] = \mathrm{P}[0,1,2,4,5,6 \text{ heads}] = \frac{44}{64}$

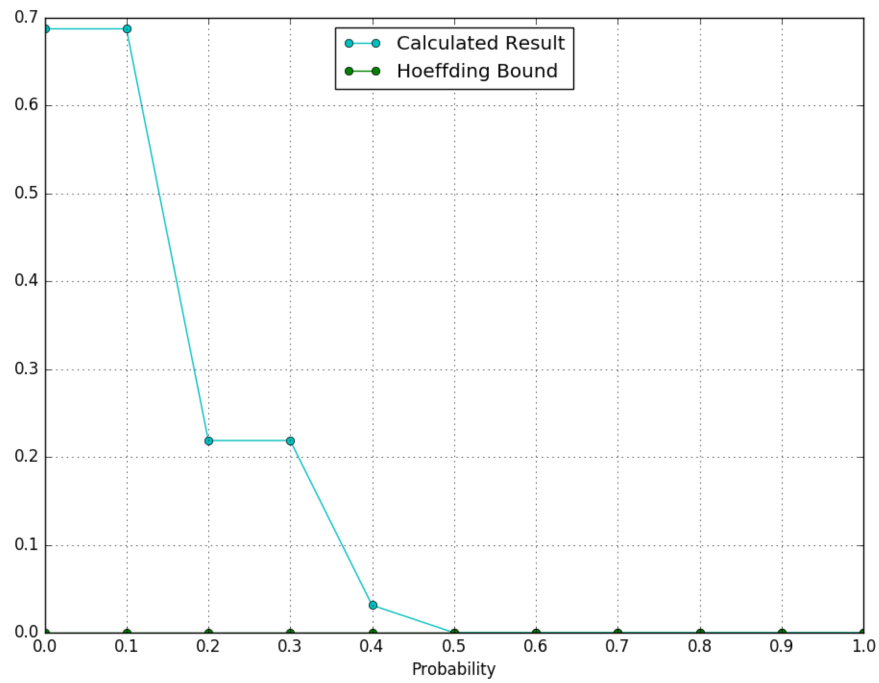
$\mathrm{P}[|\nu_i - \mu_i| > 0.2]] = \mathrm{P}[0,1,5,6 \text{ heads}] = \frac{14}{64}$

$\mathrm{P}[|\nu_i - \mu_i| > 0.3]] = \mathrm{P}[0,1,5,6 \text{ heads}] = \frac{14}{64}$

$\mathrm{P}[|\nu_i - \mu_i| > 0.4]] = \mathrm{P}[0,6 \text{ heads}] = \frac{2}{64}$

$\mathrm{P}[|\nu_i - \mu_i| > 0.5]] = 0$

The graph below shows this:



Then, to calculate $\mathrm{P}[\max|\nu_i - \mu_i| > \mathcal{E}]$, we use this:

P[A or B] = P[A] + P[B] - P[A]P[B]

For example, this calculates $\mathrm{P}[\max|\nu_i - \mu_i| > 0.0]$:

$\frac{44}{64} + \frac{44}{64} - (\frac{44}{64} * \frac{44}{64})$

Then, to calculate the Hoeffding bound, we use:

$2 * |\mathcal{H}| * e^{-2*N*\mathcal{E}^2}$

$= 2 * 2 * e^{-2*6*\mathcal{E}^2}$

$= 4 * e^{-12*\mathcal{E}^2}$

Putting these on the same graph yields this result: