CES417T - Homework 1

467261 - Yifu Wang

2018 - 09 - 07

1 LFD Problem 1.3

Follow the instruction step by step:

(a) Since $\mathbf{w}^*$ is a separation of the data. We have $y_n = \text{sign}(\mathbf{w}^{*T}\mathbf{x}_n) \neq 0$, thus $\rho = \min_{[1,N]} y_n(\mathbf{w}^{*T}\mathbf{x}_n) > 0$.

(b) By the definition of ρ and $\mathbf{w}(t+1) = \mathbf{w}(t) + y(t)\mathbf{x}(t)$, we have

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

Denote $\rho_t = \left(\mathbf{w}^T(t+1) - \mathbf{w}^T(t)\right)\mathbf{w}^*$. Then

$$\begin{aligned}
\mathbf{w}^T(t)\mathbf{w}^* &= \mathbf{w}^T(0)\mathbf{w}^* + \sum_{k=0}^{t-1}\rho_k \\
&= \sum_{k=0}^{t-1}\rho_k \\
&\geq t\rho
\end{aligned}$$

(c) Since $y(t)\mathbf{w}^T(t)\mathbf{x}(t) \leq 0$, we have

$$\begin{aligned}
\|\mathbf{w}(t)\|^2 - \|\mathbf{w}(t-1)\|^2 - \|\mathbf{x}(t-1)\|^2 &= \mathbf{w}^T(t)\mathbf{w}(t) - \mathbf{w}^T(t-1)\mathbf{w}(t-1) - x^T(t-1)x(t-1) \\
&= (\mathbf{w}(t) - \mathbf{w}(t-1))^T(\mathbf{w}(t) + \mathbf{w}(t-1)) - x^T(t-1)x(t-1) \\
&= \left(y(t-1)\mathbf{x}(t-1)\right)^T(\mathbf{w}(t) + \mathbf{w}(t-1)) - x^T(t-1)x(t-1) \\
&\leq y(t-1)\mathbf{x}^T(t-1)\mathbf{w}(t) - x^T(t-1)x(t-1) \\
&= y(t-1)\mathbf{x}^T(t-1)\left(\mathbf{w}(t) - y(t-1)\mathbf{x}(t-1)\right) \\
&= y(t-1)\mathbf{x}^T(t-1)\mathbf{w}(t-1) \\
&\leq 0
\end{aligned}$$

(d) Based on **(c)**, we have

$$\begin{aligned}
\|\mathbf{w}(t)\|^2 &= \sum_{k=1}^{t}\left(\|\mathbf{w}(k)\|^2 - \|\mathbf{w}(k-1)\|^2\right) \\
&\leq \sum_{k=1}^{t}\|\mathbf{x}(t-1)\|^2 \\
&\leq tR^2
\end{aligned}$$

namely, $\|\mathbf{w}(t)\| \leq \sqrt{t}R$.

(e) Based on **(b)** and **(d)**, we have

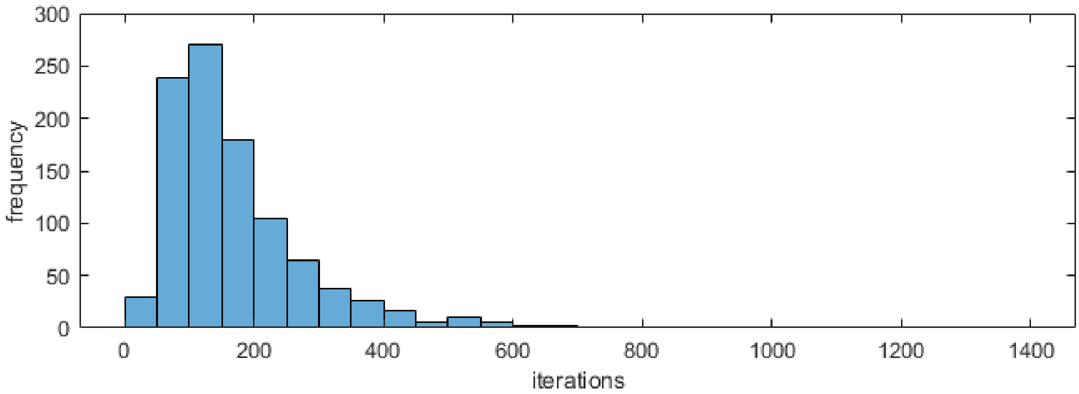
$$\frac{\mathbf{w}^T(t)\mathbf{w}^*}{\|\mathbf{w}(t)\|} \geq \frac{t\rho}{\sqrt{t}R} = \sqrt{t}\frac{\rho}{R}$$

Hence $t \leq \frac{R^2\|\mathbf{w}^*\|^2}{\rho^2}$, because $\frac{\mathbf{w}(t)}{\|\mathbf{w}(t)\|} = 1$.

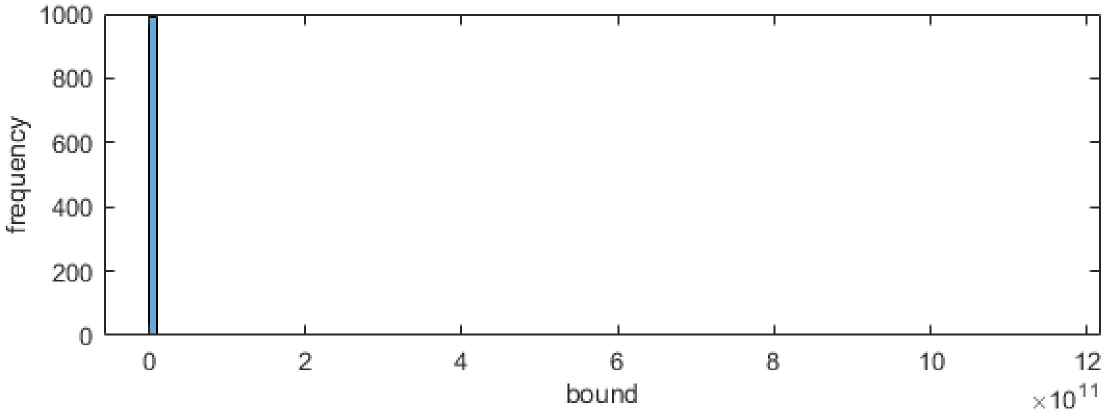
Finally, we get an upper bound of t, which means PLA will converge.

2 Lab

The histogram of iterations of 1,000 experiments seems normal:



The bound of t however is not as useful as it should be:



The average of bounds is 2.0250×10^9, which is much larger than the actual iterations. This is reasonable, the expected value of bounds is actually calculatable. But I won't do the math here, since the $\mathrm{E}\left(\frac{R^2\|\mathbf{w}^*\|^2}{\rho^2}\right)$ is extremly complex and it's not required in this assignment. To calculate this you should at least master the distribution of min value and max value and Chi square distribution. However just give a second thought you can get the idea that why this value tends to be so big, since ρ is a min value, it tend to $\mathbf{0}$, and none of R or $\|\mathbf{w}\|$ tend to $\mathbf{0}$. So I assume that the bound we calculated in problem 1 is merely a math tool to prove the convergence of PLA.

3 LED Problem 1.7

(a) Simply follow $\mathbb{P} = 1 - \left[1 - (1-\mu)^N\right]^{coins}$, we have

μ \ coins	1	1000	1000000
0.05	0.5987	1	1
0.8	1.0240×10^{-7}	1.0239×10^{-4}	0.0973

(b) Since $\mathbb{P}(k) = \begin{pmatrix} 6 \\ k \end{pmatrix} 0.5^6$, the distribution of v should be

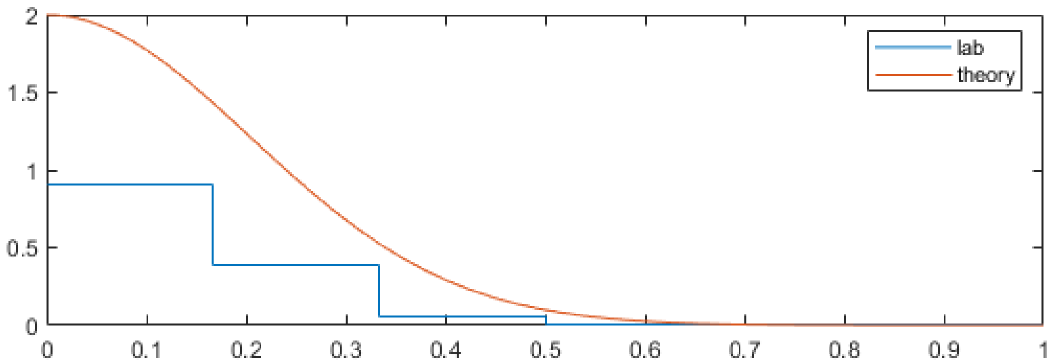
v	0	1/6	1/3	1/2	2/3	5/6	1
$P(v)$	0.0156	0.0938	0.2344	0.3125	0.2344	0.0938	0.0156

Denote $w_i = |v_i - \mu_i|$ for $i = 1, 2$. the distribution of w_i should be

w_i	0	1/6	1/3	1/2
$P(w_i)$	0.3125	0.4688	0.1876	0.0312

Then $\mathbb{P}(\max_i w_i > \epsilon) = 1 - \mathbb{P}(\max_i w_i \le \epsilon) = 1 - \mathbb{P}(w_1 \le \epsilon)\mathbb{P}(w_2 \le \epsilon)$. Thus we have

ϵ	$[0, 1/6)$	$[1/6, 1/3)$	$[1/3, 1/2)$	$[1/2, 1)$
$P(\max_i w_i > \epsilon)$	0.9023	0.3896	0.0612	0



4 LED Problem 1.8

(a) Denote the $f(x)$ as the PDF of t, we have

$$\begin{aligned} \mathbb{P}[t \ge \alpha] &= \int_\alpha^\infty f(x)dx \\ &= \frac{1}{\alpha}\int_\alpha^\infty \alpha f(x)dx \\ &\le \frac{1}{\alpha}\int_\alpha^\infty x f(x)dx \\ &= \mathbb{E}(t)/\alpha \end{aligned}$$

(b) It's obviously true based on the proof of **(a)** and $\mathbb{E}\left[(u-\mu)^2\right] = \sigma^2$.

(c) Based on **(b)**, and since u_i are iid. we show

$$
\begin{aligned}
\mathbb{E}\left[(u-\mu)^2\right] &= \mathbb{E}\left[\frac{1}{N^2}\left(\sum_{n=1}^{N}(u_n-\mu)\right)^2\right] \\
&= \frac{1}{N^2}\mathbb{E}\left[\left(\sum_{n=1}^{N}(u_n-\mu)^2+\sum_{i\neq j}(u_i-\mu)(u_j-\mu)\right)\right] \\
&= \frac{1}{N^2}\mathbb{E}\left[\sum_{n=1}^{N}(u_n-\mu)^2\right] \\
&= \frac{\sum_{n=1}^{N}\mathbb{E}(u_n-\mu)^2}{N^2} \\
&= \frac{\sigma^2}{N}
\end{aligned}
$$

Namely, $\mathbb{P}\left[(u-\mu)^2 \geq \alpha\right] \leq \frac{\sigma^2}{N\alpha}$.