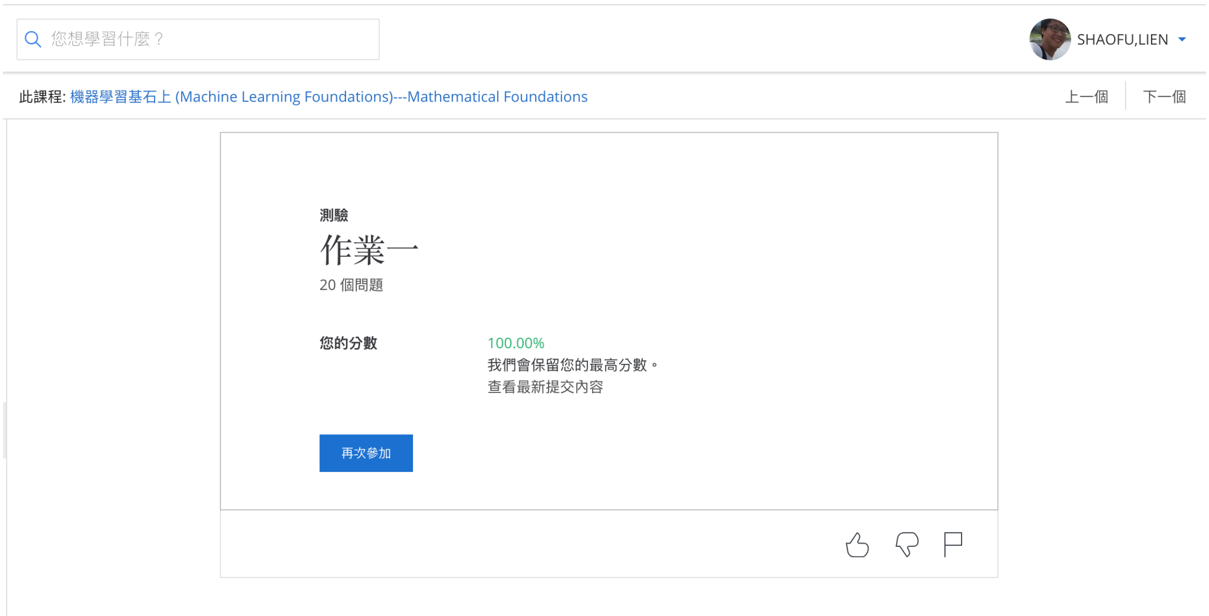
Machine Learning Foundations/hw1

連少甫/r07922107

October 12, 2018

1.



2.

根據$g(x)$及$f(x)$的特性得知$E_{OTS}(g,f)$等同於$N+1...N+L$有多少個偶數，因此$E_{OTS}(g,f)=(\left\lfloor \frac{N+L}{2} \right\rfloor - \left\lfloor \frac{N}{2} \right\rfloor)\times\frac{1}{L}$

3.

Let f_i be the set of target functions, such that $E_{OTS}(A_1(D),f)=i$, $\forall f\in f_i$, $i=0,1,2...L$

$|f_i|=\binom{L}{i}$,$E_f\{E_{OTS}(A_1(D_1),f)\}=\sum_{i=0}^{L}\frac{|f_i|}{2^L}\times i\times\frac{1}{L}=\sum_{i=0}^{L}\frac{\binom{L}{i}}{2^L}\times i\times\frac{1}{L}$.

Let g_i be the set of target functions, such that $E_{OTS}(A_2(D),g)=i$, $\forall g\in g_i$, $i=0,1,2...L$

$|g_i|=\binom{L}{i}$,$E_g\{E_{OTS}(A_1(D_2),g)\}=\sum_{i=0}^{L}\frac{|g_i|}{2^L}\times i\times\frac{1}{L}=\sum_{i=0}^{L}\frac{\binom{L}{i}}{2^L}\times i\times\frac{1}{L}$.

So that we have $E_f\{E_{OTS}(A_1(D_1),f)\}=\sum_{i=0}^{L}\frac{\binom{L}{i}}{2^L}\times i\times\frac{1}{L}=E_f\{E_{OTS}(A_1(D_2),f)\}$

4.

令$P(x=i)$代表10個抽樣裡面有i個橘球的機率，則

$P(x=0)=0.2^{10}$, $P(x=1)=10\times0.2^9\times0.8$

$P(x=10)=0.8^{10}$, $P(x=9)=10\times0.8^9\times0.2$

所以

$\nu <= 0.1$的機率為$0.2^{10} + 10 \times 0.2^9 \times 0.8 = 4.1984 \times e^{-6}$!!!!!!

$\nu >= 0.9$的機率為$0.8^{10} + 10 \times 0.8^9 \times 0.2 = 0.3758$

5.

我們可以從A或D中取出綠色的1，機率為$\frac{2}{4}$

因此五個綠色1的機率為$(\frac{2}{4})^5 = \frac{1}{32}$

6.

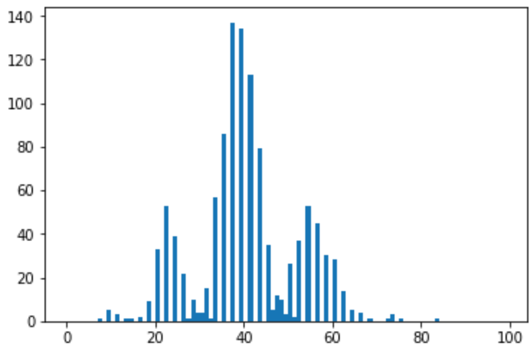
已知綠色的1從AD取出、2從BD取出、3從AD取出、4從BC取出、5從AC取出及6從BC取出，因此只需計算從AD、BD、BC、AC取出的機率再扣掉全從A、B、C、D取出的機率即可。

因此機率為，$4 \times (\frac{2}{4})^5 - 4(\frac{1}{4})^5 = \frac{31}{256}$。

7.

平均更新次數為:39.12

Histogram:



Bonus:

Prove M is exist:

$w_{t+1} \leftarrow w_t + y_{n(t)} x_{n(t)} M$, such that $y_{n(t)} w_{t+1}^T x_{n(t)} > 0$

if $y_{n(t)} > 0$ then $w_t^T x_{n(t)} < 0$

$$y_{n(t)}(w_t + y_{n(t)} x_{n(t)} M)^T x_{n(t)} > 0 \Rightarrow w_t^T x_{n(t)} + M y_{n(t)} x_{n(t)}^T x_{n(t)} > 0 \Rightarrow M > \frac{-w_t^T x_{n(t)}}{y_{n(t)} x_{n(t)}^T x_{n(t)}} > 0$$

$\Rightarrow$ M is the smallest integer so we get $M = \lfloor \frac{-w_t^T x_{n(t)}}{y_{n(t)} x_{n(t)}^T x_{n(t)}} \rfloor + 1$

if $y_{n(t)} < 0$ then $w_t^T x_{n(t)} > 0$

$$y_{n(t)}(w_t + y_{n(t)}x_{n(t)}M)^T x_{n(t)} > 0 \ \Rightarrow w_t^T x_{n(t)} + M y_{n(t)} x_{n(t)}^T x_{n(t)} < 0 \Rightarrow M > \frac{-w_t^T x_{n(t)}}{y_{n(t)} x_{n(t)}^T x_{n(t)}} > 0$$

$\Rightarrow$ M is the smallest integer so we get $M = \lfloor \frac{-w_t^T x_{n(t)}}{y_{n(t)} x_{n(t)}^T x_{n(t)}} \rfloor + 1$.

Prove the algorithm will halt with a perfect line when the data set is linear separable:

$$w_f^T w_{t+1} \geq w_f^T w_t + \min_n m_t y_n w_f^T x_n \geq w_f^T w_0 + T(m_1 + m_2 + \dots m_t) \min_n y_n w_f^T x_n$$

$$|w_{t+1}|^2 = |w_t^2| + 2w_t m_t y_{n(t)} x_{n(t)} + |m_t y_{n(t)} x_{n(t)}|^2 \leq |w_t^2| + \max_n |m_t y_{n(t)} x_{n(t)}|^2 \leq |w_0|^2 + T(m_1^2 + m_2^2 + m_3^2 \dots m_t^2) \max_n |x_n|^2$$

$$\Rightarrow \frac{w_f^T w_t}{|w_f||w_t|} \geq \frac{T(m_1 + m_2 + \dots m_t) \min_n y_n w_f^T x_n}{|w_f| \sqrt{T} \sqrt{(m_1^2 + m_2^2 + m_3^2 \dots m_t^2)} \max_n |x_n|}$$

$$0 \leq \frac{(m_1 + m_2 + \dots m_t)}{\sqrt{(m_1^2 + m_2^2 + m_3^2 \dots m_t^2)}} \leq 1 \ \text{ let } M = \frac{(m_1 + m_2 + \dots m_t)}{\sqrt{(m_1^2 + m_2^2 + m_3^2 \dots m_t^2)}}$$

$$\Rightarrow \frac{w_f^T w_t}{|w_f||w_t|} \geq \frac{\sqrt{T} \min_n y_n w_f^T x_n}{|w_f| \max_n |x_n|} \times M$$

we prove that the algorithm will halt because $\frac{\min_n y_n w_f^T x_n}{|w_f| \max_n |x_n|} \times M$ is a constant and T is the number of iteration .