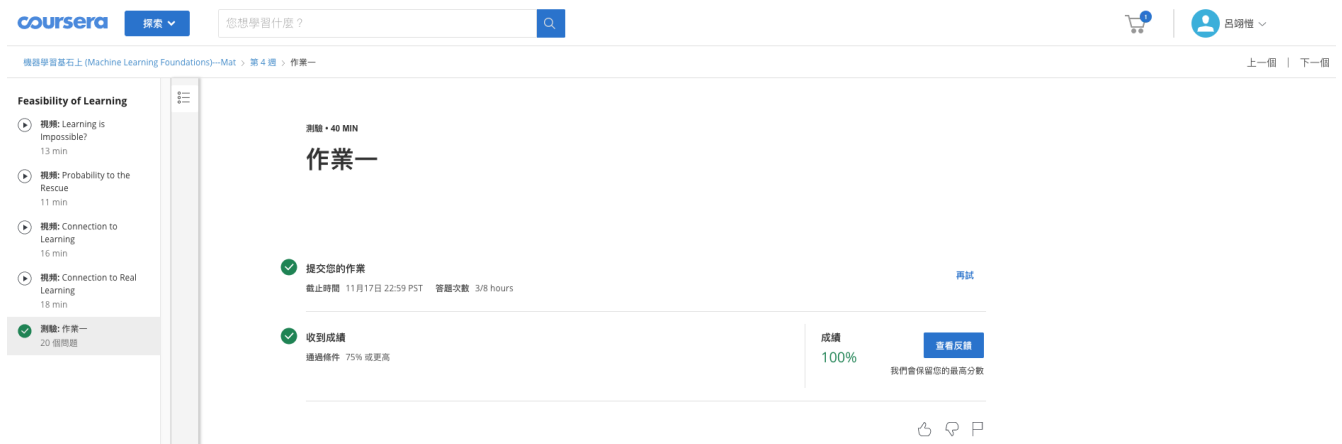


Machine Learning Foundations

Homework #1

1.



2.

semi-supervised learning可以應用在網頁分類的問題上，例如：分辨該網頁是否是垃圾網站。如果用人工分辨是否是垃圾網站，是很耗力費時的，例如Google有舉報垃圾網站的制度，會對每個遭到檢舉的網站展開調查，而用semi-supervised learning就可以用部份已知的垃圾網頁去分辨說其他網頁是不是垃圾網站。

3.

題目中說test inputs有L筆，因此代表得到的f會有 2^L 種。題目又說每一種f發生的機率相等，因此可以得知每一種f發生的機率是 $\frac{1}{2^L}$ 。

$$\text{欲求f答錯的期望值, } E_f\{E_{ots}(A(D), f)\} = \frac{1}{2^L} \frac{1}{L} \sum_{l=1}^L [g(x_{N+l}) \neq f(x_{N+l})]$$

其中 $\sum_{l=1}^L [g(x_{N+l}) \neq f(x_{N+l})]$ 為答錯的數量總和

$$\begin{aligned} \therefore E_f\{E_{ots}(A(D), f)\} &= \frac{1}{2^L} \frac{1}{L} \sum_{l=1}^L [g(x_{N+l}) \neq f(x_{N+l})] \\ &= \frac{1}{2^L} \frac{1}{L} \left(0 \binom{L}{0} + 1 \binom{L}{1} + 2 \binom{L}{2} + \dots + L \binom{L}{L} \right) \end{aligned}$$

答錯0個的
可能情況

答錯1個的
可能情況

答錯2個的
可能情況

答錯L個的
可能情況

$$= \frac{1}{2^L} \frac{1}{L} \sum_{i=0}^L i \binom{L}{i}$$

因此無論使用何種演算法，其期望值都會是上面的算法，得到一樣的期望值

4.

要得到5個綠色的1，代表5個骰子只能是A或D

1個骰子抽到A或D的機率： $\frac{1}{2}$ ，所以5個骰子抽到A或D的機率： $(\frac{1}{2})^5 = \frac{1}{32}$

5.

全部骰子的1是綠色的：5個骰子只能是A或D

全部骰子的2是綠色的：5個骰子只能是B或D

全部骰子的3是綠色的：5個骰子只能是A或D

全部骰子的4是綠色的：5個骰子只能是B或C

全部骰子的5是綠色的：5個骰子只能是A或C

全部骰子的6是綠色的：5個骰子只能是B或C

∴要得到全部骰子的某個數字是綠色的，5個骰子只能是全部是A或D、全部是B或D、全部是

B或C、全部是A或C，總共四種情況，而發生這四種情況的機率總和是 $(\frac{1}{2})^5 * 4 = \frac{1}{8}$

但這四種情況中，全部是A和全部是B和全部是C和全部是D被重複計算，因此必須扣掉五顆

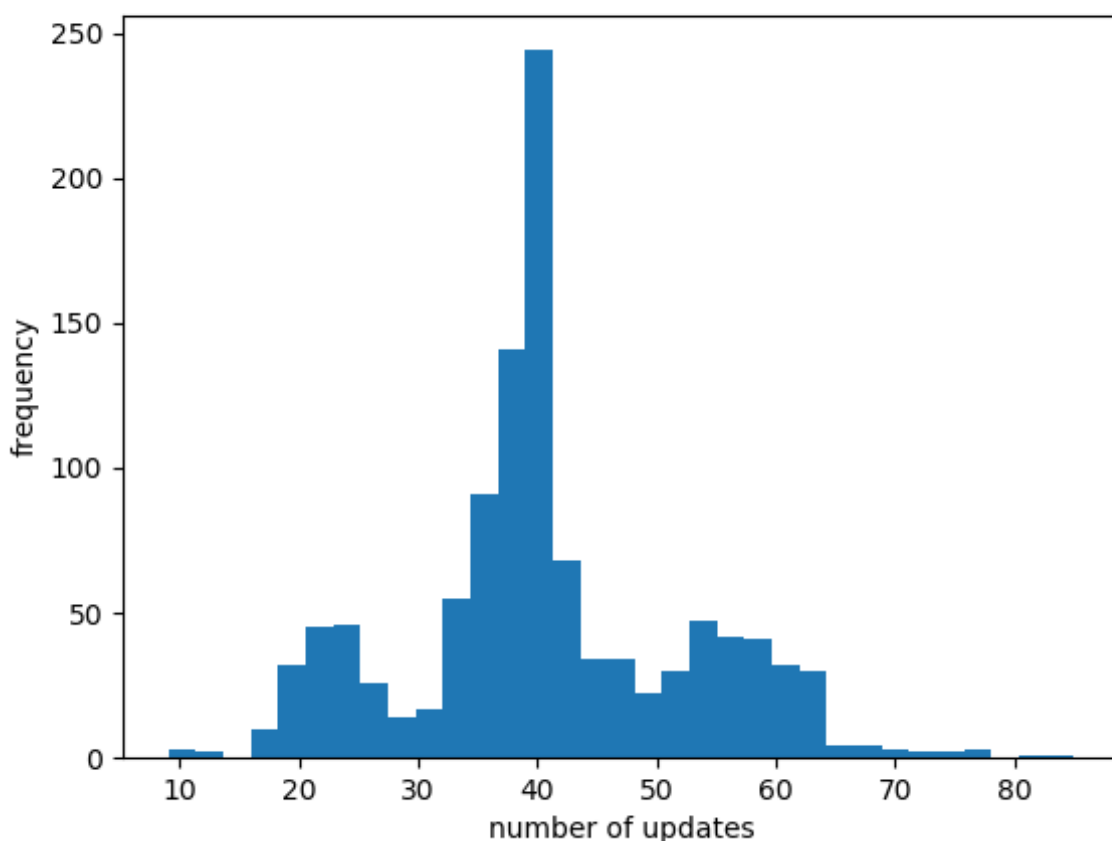
骰子全部是A和全部是B和全部是C和全部是D的機率，而該機率為 $(\frac{1}{4})^5 * 4 = \frac{1}{256}$

所以得到全部骰子的某個數字是綠色的機率為 $\frac{1}{8} - \frac{1}{256} = \frac{31}{256}$

上一題的5個骰子只能是A或D兩種骰子，但這題的骰子可以是全部是A或D、全部是B或D、全部是B或C、全部是A或C四種情況，但在計算機率的時候，必須考慮這四種情況是否有重疊的可能，而這題的四種情況是有交集的，所以還必須扣掉交集的機率，因此這是我認為這題與上一題較大的差異。

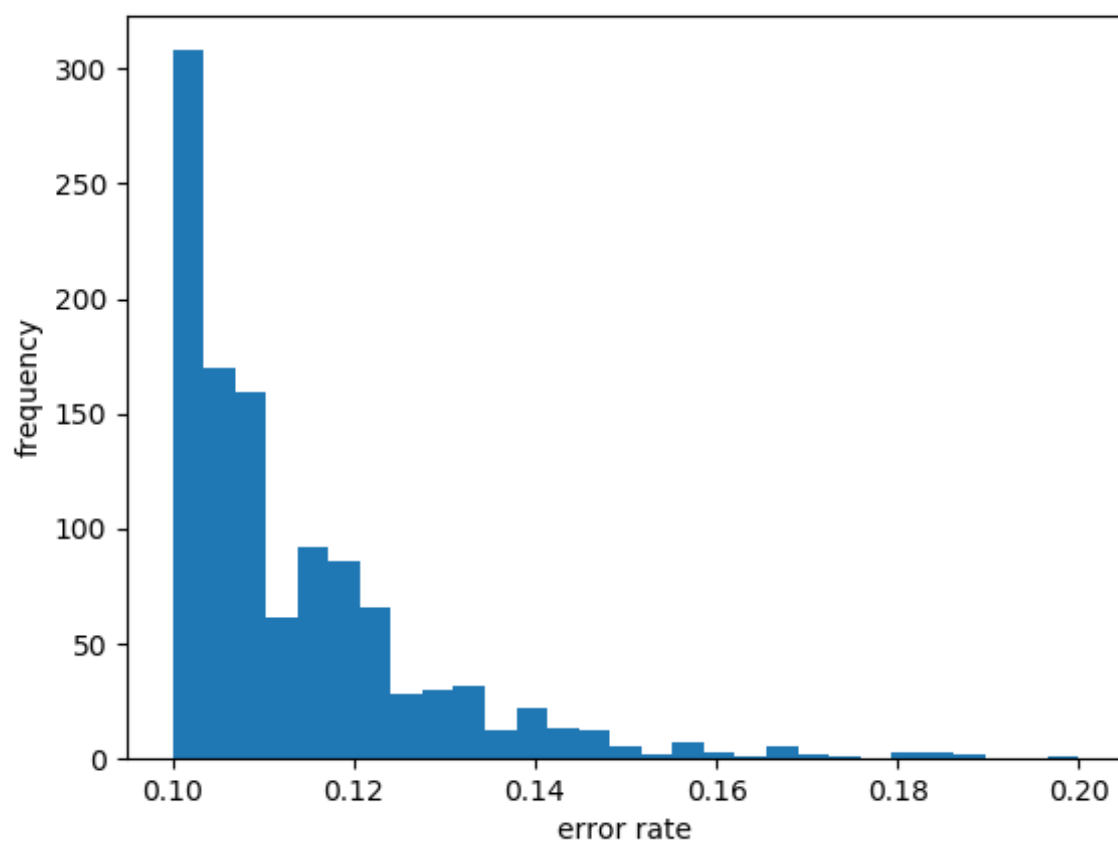
6.

average number of updates: 40.646536412078156



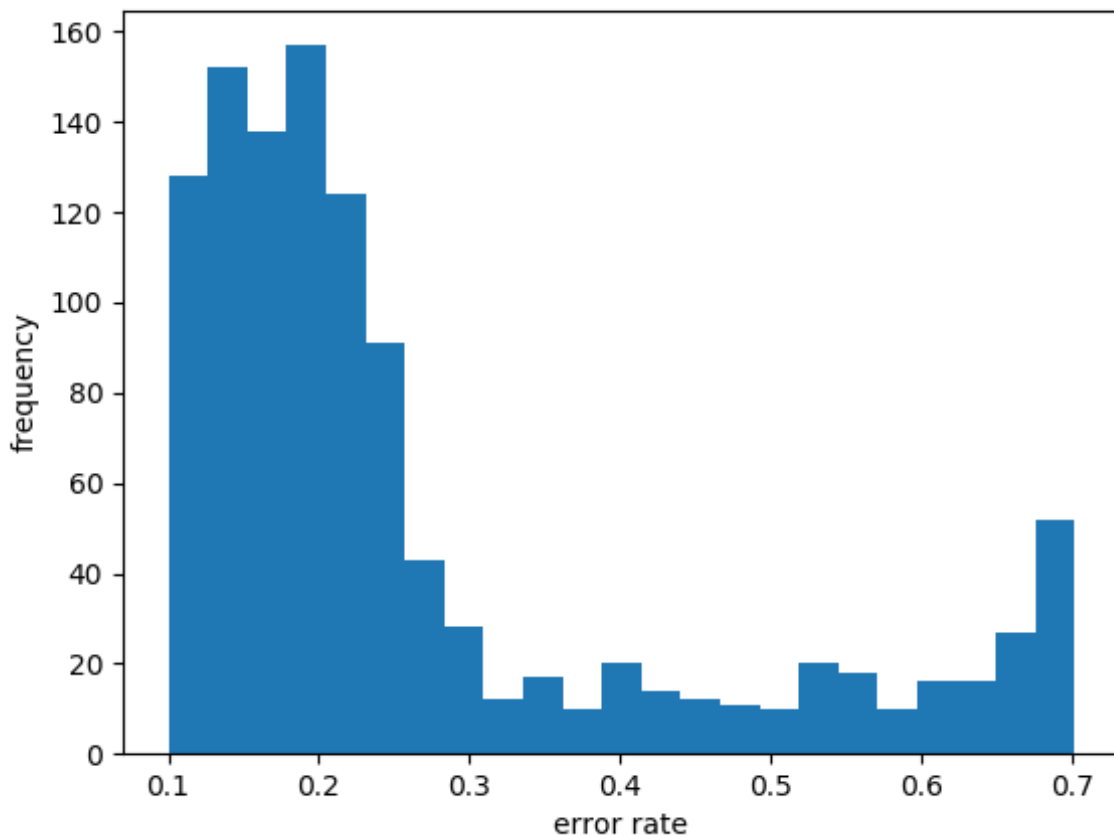
7.

average error rate: 0.113439



8.

average error rate: 0.267687



和上一題相比，這題的錯誤率比較高。此外，從兩題的histogram可以發現，上一題的error rate只有在0.1~0.2之間，但這題的error rate卻有不少次的錯誤率將近0.7。我認為是因為這題的所使用的w是更新100次後的w，所找到的w只有對當下的input能有好的預測，但對其他的input卻不一定有好的預測效果。而上一題使用的w是確定比當前的w好之後才更新，因此w只會不斷的進步，能對整體都有不錯的預測效果。

9.

Dr. Learn的計劃不會實現。

$$T \leq \frac{R^2}{\rho^2}, \text{ 其中 } R^2 = \max_n ||x_n||^2, \rho = \min_n y_n \frac{w_f^T}{||w_f||} x_n$$

$$\text{如果將 } x_n \text{ 除以 } 10, \text{ 代入上面的公式後, } T \leq \frac{\frac{1}{100} R^2}{\frac{1}{100} \rho^2} = \frac{R^2}{\rho^2},$$

可以發現T的upper bound仍然不變。