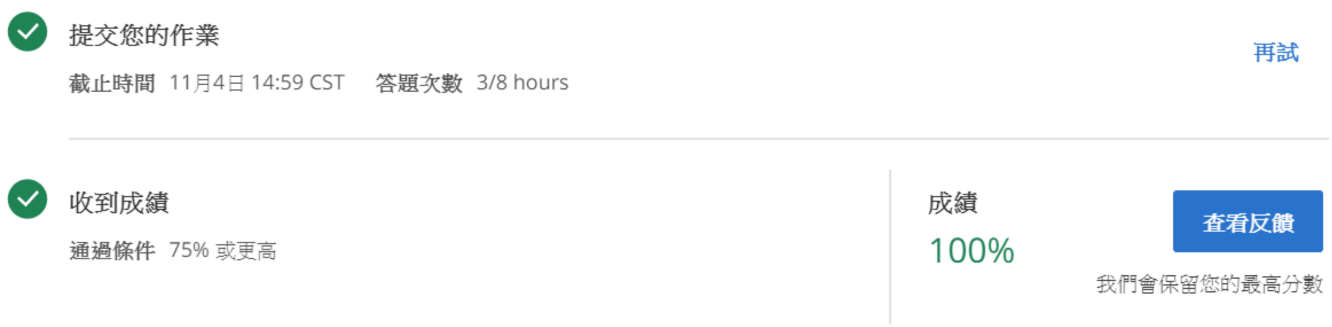
ML- HW1 資工三 蔡宥杏 b06902054

Problem 1.



Problem 2.

Semi-supervised learning 也稱為半監督學習，是指大部分的 training data 沒有 label 只有少部分才具有 label 的學習方式。在現今的聲音感測中，特別是非人聲的領域像是都市噪音種類、野外動物聲音辨識等等，本身可能已經不具有足夠的資料，有 label 的比例也不高，這時候若能利用 semi-supervised 的方式進行 learning，或許就可以提高 model 的準確性。

Problem 3.

a. We knows that：
   - A target function f can "generate" $\mathcal{D}$ in a noiseless setting if $f(x_n) = y_n$ for all $(x_n, y_n) \in \mathcal{D}$. Therefore, according to the description of the question, there is totally 2^L functions can "generate" $\mathcal{D}$ in a noiseless setting.
   - The definition of expectation of X is : $\mathrm{E[X]} = \sum_{i=1}^{k} x_i\, p_i$, $p_i = P(x = x_i)$
   - The *OTS* is : $E_{OTS}(g, f) = \frac{1}{L}\sum_{\ell=1}^{L} [\![g(\mathrm{x}_{N+\ell}) \neq f(\mathrm{x}_{N+\ell})]\!]$

b. we find out the probability under different Off-Training-Set error:
   - $\mathrm{P}\left(E_{OTS}(\mathcal{A}(\mathcal{D}), f) = \frac{k}{L}\right) = \frac{C_k^L}{2^L}$ $, when\ \#(\,[\![g(\mathrm{x}_{N+\ell}) \neq f(\mathrm{x}_{N+\ell})]\!]\,) = k$

c. Thus, we can calculate the answer:
   - $\mathbb{E}_f\{E_{OTS}(\mathcal{A}(\mathcal{D}), f)\} = \sum_{k=0}^{L} \frac{k}{L} \times \frac{C_k^L}{2^L} = \sum_{k=0}^{L} \frac{kC_k^L}{L2^L}$

$$= \sum_{k=1}^{L} \frac{LC_{k-1}^{L-1}}{L2^L} = \sum_{k=1}^{L} \frac{C_{k-1}^{L-1}}{2^L} = \frac{2^{L-1}}{2^L} = \frac{1}{2} = constant$$

Problem 4.

取 5 個骰子，求得 5 個都有綠色的 1 的機率是多少？根據 ABCD 四種骰子，可知每個骰子：

有個綠色的 1，Type A 或 Type D

沒有綠色的 1，Type B 或 Type C

因此若要 5 個骰子都有綠色的 1，代表取出的 5 個骰子都必須是 Type A 或 Type D。因此機率為：

$$p = \frac{2^5}{4^5} = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

Problem 5.

取 5 個骰子，求某些數字都是綠色的機率是多少？分別觀察 1 到 6：

1 : Type A or Type D	4 : Type B or Type C
2 : Type B or Type D	5 : Type A or Type C
3 : Type A or Type D	6 : Type B or Type C

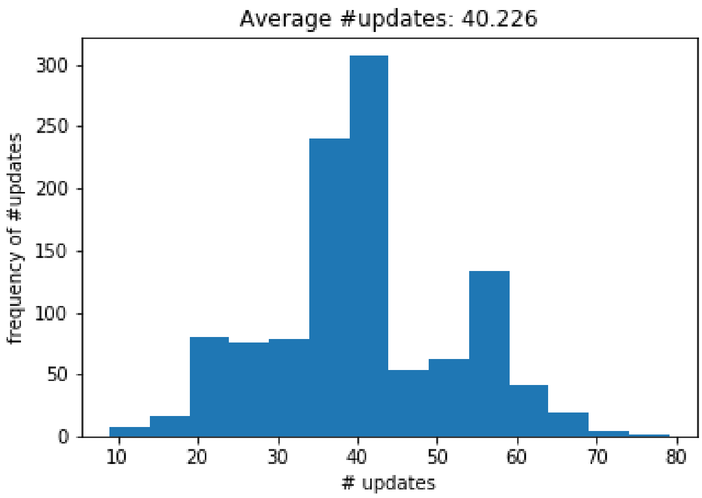
其中數字 1 和 3、4 和 6 會分別同時成立，因此總共有四種組合(A,D) (B,C) (B,D) (A,C)，不過需要在另外扣掉全部都為 A,B,C,D 重複計算的 4 次，例如：在(A,D)的組合中會有 AAAAA 或 DDDDD，各自在(A,C)或(B,D)的組合中就會被重複算到一次。因此機率為：

$$p = \frac{2^5 \times 4 - 4}{4^5} = \frac{31}{256}$$

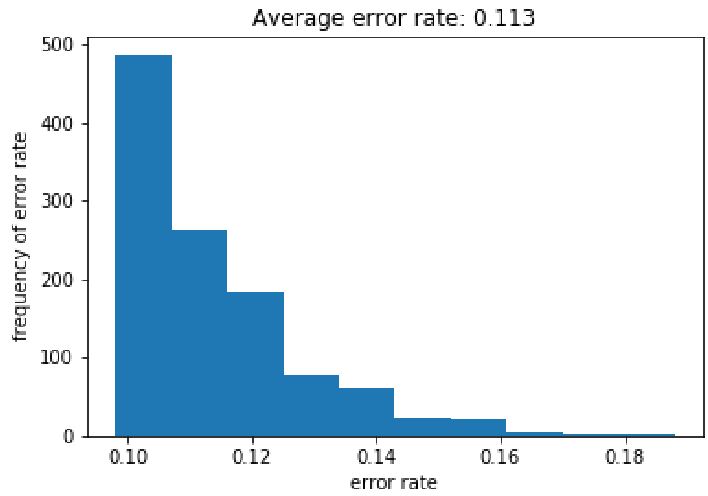
與 problem4 進行比較，從 problem5 的計算公式及題意就可以知道這算是 problem4 的延伸，前者只討論單一數字 1 所以只有組合(A,D)，後者則是同時考慮所有的數字所以才會歸納出(A,D) (B,C) (B,D) (A,C)這四種組合，再扣除重複計算的部分就可以得到答案，因此計算也可以看成由 problem4 的答案推算，如以下：

$$p = \frac{1}{32} \times 4 - \frac{1}{4^5} \times 4 = \frac{31}{256}$$

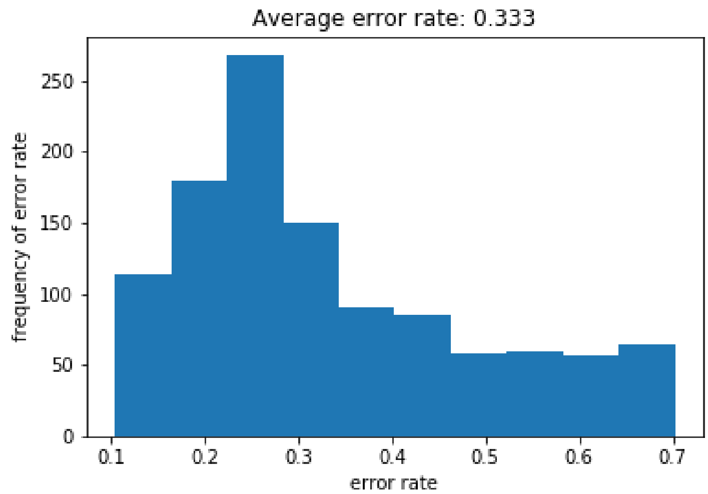
Problem 6. PLA + different random seed + 1126 times



Problem 7. PLA + 100 updates + W_pocket + different random seed + 1126 times



Problem 8. PLA + 100 updates + W_100 + different random seed + 1126 times



Compare the result of problem7 and problem8, we can find that the error rate of the algorithm which use w_hat to remember the best one so far is lower than another one. As a result, we can know that if the test data is not linear separable, only use the basic PLA algorithm can't let us get a good enough result. Using the PLA and Pocket Algorithm together can reduce the error rate effectively.

Problem7 – PLA and Pocket algorithm, Average error rate = 0.113
Problem8 – Only PLA algorithm, Average error rate = 0.333

Problem 9.

This plan will not work, the algorithm wound not run 10 times faster.

The upper-bound of the number of correction T is $\mathrm{T} \leq \frac{R^2}{\rho^2}$ ($R^2 = \max_n \|x_n\|^2$, $\rho = \min_n y_n \frac{{W_f}^T}{\|W_f\|} x_n$) , if we divide x_n with 10, then R^2 and ρ^2 will be divided with 100 at the same time, there is no effect on the upper-bound T.