

機器學習基石 Homework#1

1. Go register for the Coursera version of the first part of the class and solve its homework 1.



2. Describe an application of semi-supervised learning within 10 English or Chinese sentences.

語音識別使用半監督式學習提取重要訊息

在語音辨識領域中希望 label 的標註有較高的正確性，但標註的時間與人力成本也相對較高，因此使用半監督式學習讓模型擁有趨近於監督式學習的分類能力，同時也減少許多相對於監督式學習中標註的成本

3. please prove or disprove

Prove : $E_f \{ E_{OTs}(A(D), f) \} = \text{constant}$

本題的解題關鍵是”For any deterministic algorithm A, if all those f that can “generate” D in a noiseless setting are equally likely in probability.”

亦即對於所有的演算法 A，目標函數 f 都可以擬合出 D 且每種狀況的機率相等，我們可以知道如果 f 要擬合 A(D)共有 N 種不同的可能，因此 f 擬合出正確答案的機率便只有 1/N，所以 E_{OTs} 的期望值會是定值

4. If we pick 5 dice from the bag, what is the probability that we get five green 1's?

綠色的 1 只有 A 跟 D 兩個 class，因此能把題目改寫成抽到五個(A 或 D)，在四個 class 的骰子數量一樣的情況下，抽到(A 或 D)的機率就是 1/2，那麼要抽到五個(A 或 D)的機率就是：

$$(1/2)^5 = 1/32$$

5. If we pick 5 dice from the bag, what is the probability that we get "some number" that is purely green?

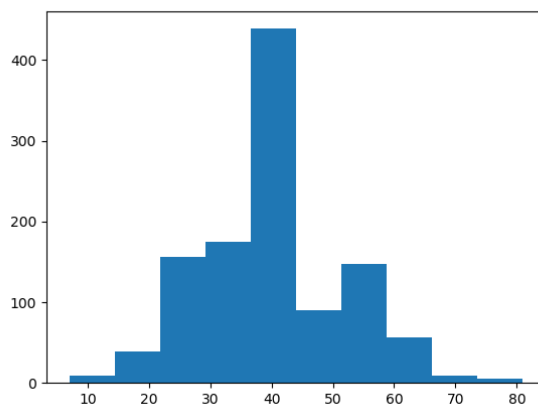
綠色的 1~6 對應的 class 分別為(A 或 D), (B 或 D), (A 或 D), (B 或 C), (A 或 C), (B 或 C)，因此只要計算(A 或 D), (A 或 C), (B 或 D), (B 或 C)聯集的機率，就是四個項的總和 $(2^5)*4$ 減去四個項的交集，全 A、全 B、全 C、全 D 共 4 種：

$$[(2^5)*4-4]/(4^5) = 31/256$$

上一題的答案經過擴分後變為 $8/256$ ，可以發現數字同顏色的題目在不指定數字的情況機率是指定數字的四倍不到，仔細觀察會發現只要全為綠色 1 就會全為綠色 3，反之亦然，同理可套用在綠色 4 跟 6

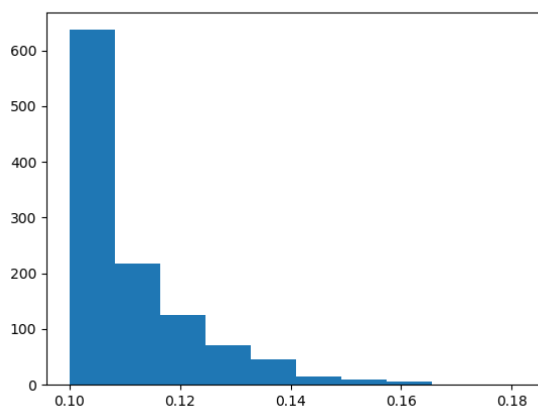
6. Please repeat your experiment for 1126 times, each with a different random seed. What is the average number of updates before the algorithm halts?

大約 40 次



7. Please repeat your experiment for 1126 times, each with a different random seed. What is the average error rate on the test set?

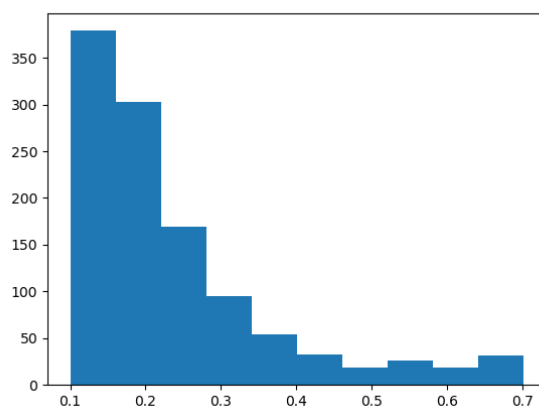
大約 0.11



8. Please repeat your experiment for 1126 times, each with a different random seed. What is the average error rate on the test set?

大約 0.238

跟上一題比較起來使用 PLA 的 error 增高了許多，仔細觀察會知道 PLA 演算法有錯誤就更正，但更正後的那條線錯誤說不定比原本的還多，所以 PLA 不一定會往 loss 最低的地方收斂，只能保證在線性可分的資料中能找出解。但使用 pocket 演算法每次找到錯誤就要量測新的線錯誤是否比舊的線多，使得整體的計算非常耗時，然而可以確定的是使用 pocket 演算法 loss 是有逐步下降的



9. Will his plan work? Why or why not?

This plan will not work.

講義中更正次數上限 T 不只取決於資料半徑 R ，還有取決於 ρ^2 項，因此我們知道如果 input data 全部縮小 10 倍的話， R^2 會縮小 100 倍，但 ρ^2 也會縮小 100 倍，兩者相除後始得更正次數上限不變，整體收斂速度也不會改變