

Machine Learning - HW#1

The Learning Problem

1. (d)

根據過去的顧客購買資料，以過去購買紀錄、瀏覽次數、網頁停留時間.....等參數作為 \mathbf{x} 、有無購買作為 y ，建立一個 binary classification 的機器學習模型，找到 $g(\mathbf{x})$ 。

Modifications of PLA

2. (e)

將原來的PLA更新式同乘以 $y_{n(t)}x_{n(t)}$ ，可以得到：

$$y_{n(t)}w_{t+1}^T x_{n(t)} = y_{n(t)}w_t^T x_{n(t)} + \|x_{n(t)}\|^2$$

若要能夠確保左式中的 $y_{n(t)}w_{t+1}^T x_{n(t)}$ 會大於零，則右式中的 $y_{n(t)}w_t^T x_{n(t)}$ 之絕對值必須小於

$$\|x_{n(t)}\|^2 \quad (y_{n(t)}w_t^T x_{n(t)} \text{ 必為負}) \quad , \text{ 而將 learning rate 設為 (E) 選項中的 } \left\lfloor \frac{-y_{n(t)}w_t^T x_{n(t)}}{\|x_{n(t)}\|^2} + 1 \right\rfloor \text{ 可}$$

以確保右式中的第二項會大於第一項的絕對值（也就是 $-y_{n(t)}w_{t+1}^T x_{n(t)}$ ），即：

$$\begin{aligned} -y_{n(t)}w_t^T x_{n(t)} &< \|x_{n(t)}\|^2 \\ \Rightarrow y_{n(t)}w_{t+1}^T x_{n(t)} &> 0 . \end{aligned}$$

3.

Extensions of PLA

4.

5. (b)

根據題目中敘述的 multi-class classification 方法，其在一個平面空間中的幾何意義可以理解為多個指向分類 k 的向量 w_k （因為內積愈大的意義即為兩向量在方向上愈靠近）。

因此，在找到一個預測錯誤的資料點 x_n （其 true label 為 y 、預測錯誤的 label 為 y' ）時，我們可以將 w 的迭代規則想像成將 w_y 轉向該資料點的方向、並將 $w_{y'}$ 轉離該資料點的方向，透過不斷的迭代，使得最終所有的 w_k^* 向量皆指向分類 k 的方向。

在 $K=2$ 、 $y_n \in \{1,2\}$ 時，可以在迭代後得到 w_1^* 、 w_2^* 兩個最終的權重向量，並根據 $\tilde{y}_n = 2y_n - 3$ 的轉換式，得知 w_1^* 向量應指向 $\tilde{y}_n = -1$ 之分類群組的方向（與 W_{PLA} 的意義相反）；同時 w_2^* 向量應指向 $\tilde{y}_n = +1$ 之分類群組的方向（與 W_{PLA} 的意義相同）。以上說明之示意圖如下。

放圖

The Learning Problems

6. (d)

由於本題中的資料並沒有標記 label（沒有 y ），且存在一個由機器學習模型自己定義的目標（self-defined goal），因此屬於 self-supervised learning。

7. (c)

在本題中，一篇文章可能屬於多個 label，因此屬於 multi-label classification。另外，整個 data set 中有部分資料有經過 label、部分資料沒有經過 label，因此屬於 semi-supervised learning。

Off-Training-Set Error

8. (b)

首先將資料的 X 定義為 (x_1, x_2) 。

Smallest $E_{ots}(g)$ ：

選擇 $(-2, 2)$ 、 $(0, 0)$ 、 $(0, 3)$ 作為 D ，經學習後的 $g(X)$ 如下：

$$g : \begin{cases} \text{if } x_1 \leq 0 : y = +1 \\ \text{else } : y = -1 \end{cases}$$

此時 U 中所有的資料皆符合 $g(X) = y$ ，因此有最小的 $E_{ots}(g) = 0$ 。

Largest $E_{ots}(g)$ ：

選擇 $(0, 0)$ 、 $(2, 1)$ 、 $(3, 0)$ 作為 D ，經學習後的 $g(X)$ 如下：

$$g : \begin{cases} \text{if } x_1 = x_2 : y = +1 \\ \text{else } : y = -1 \end{cases}$$

此時 U 中其他三筆資料皆不符合 $g(X) = y$ ，因此有最大的 $E_{ots}(g) = 1$ 。

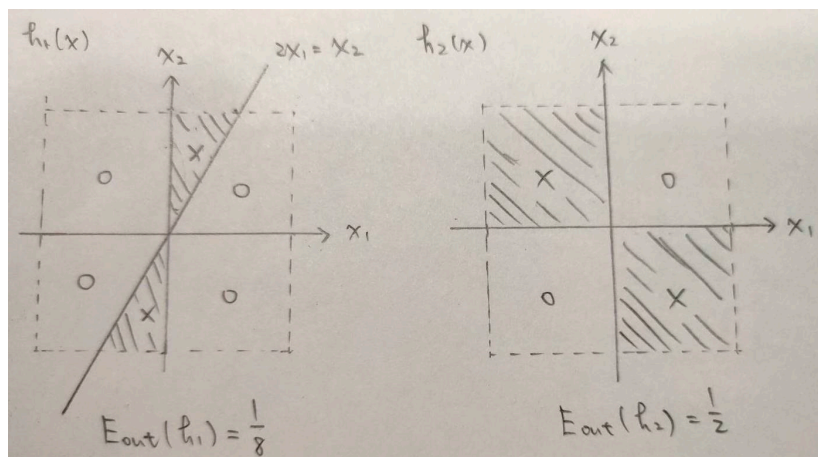
Point Estimation

9.

Bad Data

10. (a)

下圖中分別為 $h_1(x)$ 和 $h_2(x)$ 在空間中的示意圖，其中標示「O」的地方代表 $h(x) = f(x)$ ，反之則標示「X」（斜線處）。根據示意圖，可得出 $(E_{out}(h_1), E_{out}(h_2)) = (\frac{1}{8}, \frac{1}{2})$ 。



11. (b)

放圖

上圖中，標記 A、B、C 的區域分別代表以下意義：

$$A : h_1(x) = h_2(x) \quad P(A) = \frac{3}{8}$$

$$B : f = h_1 \neq h_2 \quad P(B) = \frac{4}{8}$$

$$C : f = h_2 \neq h_1 \quad P(C) = \frac{1}{8}$$

隨機產生 4 個資料點，得到 $E_{in}(h_2) = E_{in}(h_1)$ 的可能性及其機率如下：

$$4A : P(4A) = \left(\frac{3}{8}\right)^4 = \frac{81}{4096}$$

$$2A+1B+1C : P(2A + 1B + 1C) = \left(\frac{3}{8}\right)^2 \frac{4}{8} \frac{1}{8} \frac{4!}{2!} = \frac{432}{4096}$$

$$2B+2C : P(2B + 2C) = \left(\frac{4}{8}\right)^2 \left(\frac{1}{8}\right)^2 \frac{4!}{2!2!} = \frac{96}{4096}$$

$$\Rightarrow P(E_{in}(h_2) = E_{in}(h_1)) = P(4A) + P(2A + 1B + 1C) + P(2B + 2C) = \frac{609}{4096}$$

Multiple-Bin Sampling

12. (b)

放表格

P (有數字全都是綠的)

= 1 - P (每個數字都有橘的)

= 1 - P (至少一個C \cap 至少一個D)

= P (沒有C \cup 沒有D)

= P (沒有C) + P (沒有D) - P (沒有C \cap 沒有D)

$$= \left(\frac{3}{4}\right)^5 + \left(\frac{3}{4}\right)^5 - \left(\frac{2}{4}\right)^5 = \frac{454}{1024}$$

	Green	Orange
A	2, 4, 6	1, 3, 5
B	1, 2, 6	3, 4, 5
C	6	1, 2, 3, 4, 5
D	2, 3, 5	1, 4, 6

Experiment with Perceptron Learning Algorithm

13. (A)

14. (B)

15.

16.