

Deep Learning Homework 1

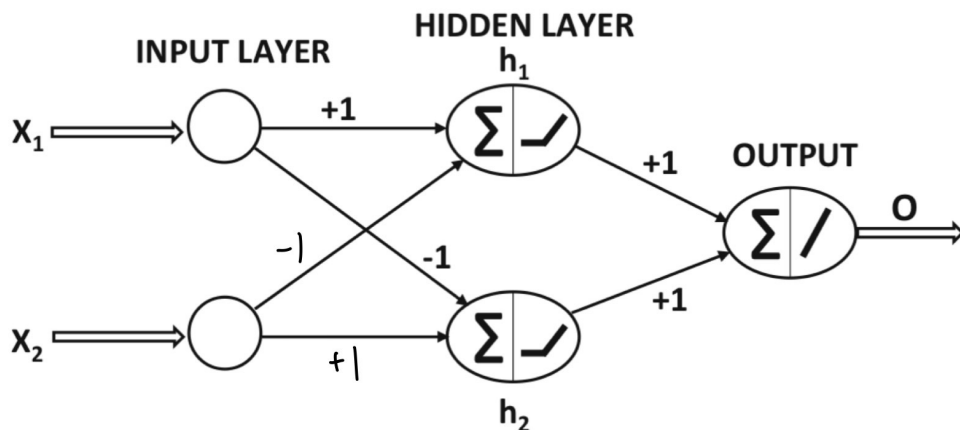
110024516 統研所 邱繼賢

Exercise 1.

建構神經網路圖如下，一層 input layer，一層 hidden layer 使用 ReLU activation function，一層 output layer 使用 identity activation function，各路徑權重如下圖所示，四個資料點經過此神經網路後即可區分成兩組：

$$\{(0,0), (1,1)\} \rightarrow \{0,0\}$$

$$\{(1,0), (0,1)\} \rightarrow \{1,1\}$$



Exercise 4.

(a) 我們的目標函數數值並不會隨著時間而變小，因此演算法也不會收斂。

(b) 因為 perceptron 是一種線性的分類器，但是本題中的兩組數據點 not linearly separable，故出現 (a) 中之現象。

Exercise 5.

$$\begin{aligned} f : z &= x_1 \cdot x_2 \\ \Rightarrow \{(-1, -1), (1, 1)\} &\xrightarrow{f} \{1, 1\} \\ \{(1, -1), (-1, 1)\} &\xrightarrow{f} \{-1, -1\} \end{aligned}$$

將 (x_1, x_2) 轉換成 z 後，兩組資料點便可以 linearly separate。

原本的 (x_1, x_2) 是 not linearly separable，但是經過了 nonlinear transformation f 將 (x_1, x_2) 映射到一維 z 後，我們就可以輕易地進行 linearly separate，由此可知，我們可以藉由 nonlinear transformation 將原本線性不可分割的資料點映射成可以線性分割的資料點，進而幫助我們進行分類。

Exercise 11.

(a) Loss function :

$$L_i = \max \{-y_i (\bar{W} \cdot \bar{X}_i), 0\}$$

Update rule of weights :

$$\bar{W} \Leftarrow \bar{W} + \alpha (y_i - \hat{y}_i) \bar{X}_i$$

$$\text{or } \bar{W} \Leftarrow \bar{W} + \alpha y_i I[-y_i (\bar{W} \cdot \bar{X}_i) > 0] \bar{X}_i$$

Accuracy = 92.7%

(b) Loss function :

$$L_i = \max \{1 - y_i (\bar{W} \cdot \bar{X}_i), 0\}$$

Update rule of weights :

$$\bar{W} \Leftarrow \bar{W} + \alpha y_i I[1 - y_i (\bar{W} \cdot \bar{X}_i) > 0] \bar{X}_i$$

Accuracy = 94.1%

(c) 使用 hinge-loss 時分類的表現更好，因為 hinge-loss 除了原本那些分類錯誤的資料點會更新 weights，連距離邊界太過接近的資料點發生時也會去更新 weights，進而導致其在預測未知的 test data 時的表現會更好。

(d) 因為 (c) 的理由，利用 hinge-loss 所訓練出的模型要確保 training data 距離邊界足夠遠，這會使得 training model 比較穩定，較不易受 training data 的改變而影響。