

1. 在感知器模型中，輸入  $x$  和权重  $w$  相乘後相加，是一種直觀且有效的計算方式。

• 數學直覺： $z = w \cdot x$  實際上是向量內積的表現形式，能夠有效表示輸入向量和权重向量的相關性。

當輸入  $x$  和权重  $w$  越相似時，它們的內積值越大，代表該輸入組合的影響也越大。

• 生活觀察：如果一個輸入的权重較高，意味著該輸入的決策過程中更為重要，會對最終的輸出值產生更大的影響。這反映了我們在生活中對不同因素進行加權平均的做法：當面對多個決策依據（如價格、品牌、品質）時，我們會給予一些重大因素更大的权重，而這些因素會對最終結果起決定性作用。

2. 1) • Perceptron: 對於多輸入單層感知器，輸入  $x_1, x_2, \dots, x_n$ ，权重  $w_1, w_2, \dots, w_n$ ，則總輸入 net input 為：

$$z = \sum_{i=1}^n w_i x_i = w_1 x_1 + w_2 x_2 + \dots + w_n x_n$$

輸出則是將  $z$  通過 sigmoid function:

$$y = \sigma(z) = \frac{1}{1 + e^{-z}}$$

• Logistic regression: 在給定輸入  $x$  的條件下，事件  $y=1$  發生的機率：

$$P(y=1|x) = \frac{1}{1 + e^{-(w_1 x_1 + w_2 x_2 + \dots + w_n x_n)}}$$

⇒ 兩者的數學公式完全一致，因此他們是等價的。

2) Delta Rule:

$$\Delta w_i = -\alpha \frac{\partial E}{\partial w_i}, \quad \alpha \text{ 為學習率}$$

$$\text{假設 Loss function} = \frac{1}{2}(y - \hat{y})^2, \quad \hat{y} = \sigma(z), \quad z = \sum w_i x_i,$$

則透過 chain rule:

$$\Delta w_i = -\alpha \frac{\partial E}{\partial w_i} = -\alpha \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial w_i}$$

其中,

$$\frac{\partial E}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}} \left[ \frac{1}{2}(y - \hat{y})^2 \right] = -(y - \hat{y})$$

$$\frac{\partial \hat{y}}{\partial z} = \hat{y}(1 - \hat{y}) \rightarrow \text{sigmoid 的導數}$$

$$\frac{\partial z}{\partial w_i} = x_i$$

$$\Rightarrow \Delta w_i = -\alpha [-(y - \hat{y})] \hat{y}(1 - \hat{y}) x_i = \alpha (y - \hat{y}) \hat{y}(1 - \hat{y}) x_i$$

error ↑
input ↑  
↓ learning rate
↓ sigmoid derivation

- 3) • Perceptron : 感知器使用 sigmoid function, 輸出介於  $0 \sim 1$ , 但本身並沒有明確地定義這輸出是機率值, 而是作為分類的依據。
- Logistic regression: 在邏輯迴歸中, 輸出明確解釋為機率, 即某事件發生的機率, 並進行推理, 這是邏輯迴歸的核心。

3.

(1) 我們的目標是根據個體的身高（單位：公分）和體重（單位：公斤）預測其性別。這是一個二元分類問題，性別變量有兩個可能的類別：Male 和 Female。我們將使用多層感知器（MLP）來進行模型訓練，並根據不同的模型參數來優化預測性能。資料來源：<https://www.kaggle.com/datasets/saranpannasuriyaporn/male-female-height-and-weight?resource=download>

(2) 以下為部分資料，樣本數共有3000人，男女比例各半。

Height	Weight	Sex
165.65	35.41	Female
148.53	74.45	Female
167.04	81.22	Male
161.54	71.47	Male
174.31	78.18	Male

(3) 由於 MLP Playground 只能接受兩個變量進行訓練，我們可以選擇使用 "Height" 和 "Weight" 這兩個特徵來進行預測性別（"Sex"）的分類問題，不需要進行複雜的特徵工程。

(4) Ablation Study

Training loss	Test loss	noise	ratio of training to test data	batch size	learning rate	activation function	number of hidden layers	number of neurons per hidden layer	neurons
0.189	0.189	0	70%	10	0.03	sigmoid	1		4 x1,x2
0.189	0.19	0	80%	10	0.03	sigmoid	1		4 x1,x2
0.189	0.19	0	70%	20	0.03	sigmoid	1		4 x1,x2
0.19	0.19	0	70%	30	0.03	sigmoid	1		4 x1,x2
0.188	0.19	0	70%	10	0.1	sigmoid	1		4 x1,x2
0.191	0.19	0	70%	10	0.01	sigmoid	1		4 x1,x2
0.256	0.263	0	70%	10	0.001	sigmoid	1		4 x1,x2
0.189	0.19	0	70%	10	0.03	Tanh	1		4 x1,x2
0.188	0.191	0	70%	10	0.03	ReLU	1		4 x1,x2
0.192	0.191	0	70%	10	0.03	sigmoid	0		4 x1,x2
0.189	0.189	0	70%	10	0.03	sigmoid	2		4 x1,x2
0.188	0.189	0	70%	10	0.03	sigmoid	3		4 x1,x2
0.5	0.5	0	70%	10	0.03	sigmoid	4		4 x1,x2
0.192	0.192	0	70%	10	0.03	sigmoid	1		1 x1,x2
0.188	0.19	0	70%	10	0.03	sigmoid	1		2 x1,x2
0.189	0.19	0	70%	10	0.03	sigmoid	1		3 x1,x2
0.188	0.19	0	70%	10	0.03	sigmoid	1		5 x1,x2
0.187	0.19	0	70%	10	0.03	sigmoid	1		6 x1,x2
0.188	0.191	0	70%	10	0.03	sigmoid	1		4 x1,x2,x1x2
0.187	0.19	0	70%	10	0.03	sigmoid	1		4 x1,x2,x1x2,x1^2
0.187	0.194	0	70%	10	0.03	sigmoid	1		4 x1,x2,x1x2,x1^2,x2^2
0.478	0.477	0	70%	10	0.03	sigmoid	1		4 x1^2,x2^2

(5) Analysis

從表中可以看到，一開始先全部使用預設值，得出結果Training loss=0.189, Test loss=0.189。接著開始調配各種不同參數的設定。

- noise：原本想調整這個參數，但是這個鍵似乎按不了。
- ratio of training to test data：觀察7:3和8:2的比例，發現切7:3的結果較佳，因此往後皆使用7:3的比例進行實驗。
- batch size：觀察10,20,30的數量，發現當數量為10時結果最佳，因此往後也都使用這個配置。
- learning rate：觀察0.1,0.03,0.01,0.001的學習率，可以看到學習率為0.001的結果最差，因為我設定大概epoch皆跑500個epoch，在這個迭代次數下，學習率過低模型還沒有收斂。其餘雖然結果皆差不多，但還是保持使用原始預設值learning rate=0.03。

- activation function：觀察sigmoid,Tanh,ReLU三種激活函數，發現結果也都差不多，因此保持使用原始預設值sigmoid。
- number of hidden layers：分別進行0~4層隱藏層，可以發現到4層隱藏層時結果非常差，可能是參數過多，簡單的任務不需要這麼多的隱藏層，因此往後設定皆使用1層隱藏層。
- number of neurons per hidden layer：分別進行1~6個的隱藏層神經元個數，可以發現結果依然皆差不多，因此維持使用預設值4個。
- neurons：從表中可以發現當使用二次方時，結果非常差，可以得知此種資料並不適合將其平方，因此依然選擇最簡單的x1,x2兩種neurons。

因此從結果最佳和最簡單的模型為挑選原則後，最終模型參數設置如下：

- noise：0
- ratio of training to test data：7:3
- batch size：10
- learning rate：0.03
- activation function：sigmoid
- number of hidden layers：1
- number of neurons per hidden layer：4
- neurons：x1,x2

最後再重新生成資料幾次，得到最佳的實驗結果如下圖所示，Test loss=0.184。

