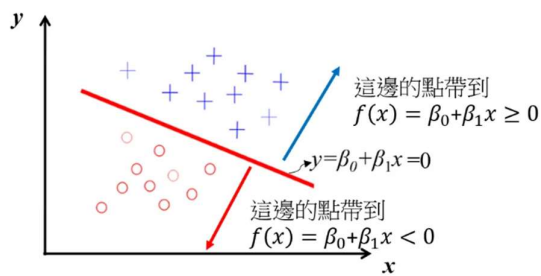
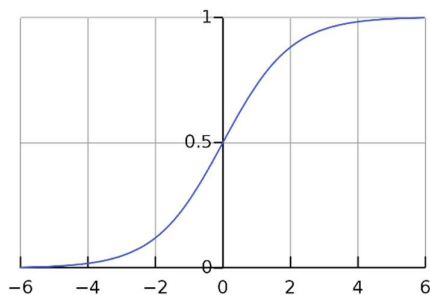


邏輯回歸:一種線性回歸的推廣，專門用於解決二分類問題。

Sigmoid 函數: $f_w(x) = \frac{1}{1+e^{-f(x)}} = \sigma(x)$, $f(x) = \mathbf{w}x + b$

特點:會把所有輸入值轉成 0 和 1 之間



單一樣本出現機率:

$$P(y = 1|\mathbf{x}) = f_w(\mathbf{x}) = \frac{1}{1+e^{-f(\mathbf{x})}}(x) = \sigma(\mathbf{x}\mathbf{w})$$

m 個樣本出現機率:

$$\prod_{i=1}^m (f_w(\mathbf{x}_i)^{y_i} (1 - f_w(\mathbf{x}_i))^{1-y_i})$$

代價函數:取 log

$$L(\mathbf{w}) = -\frac{1}{m} \sum_{i=1}^m (y_i \log(f_w(\mathbf{x}_i)) + (1 - y_i) \log(1 - f_w(\mathbf{x}_i)))$$

交叉熵:

$$L(\mathbf{w}) = -\frac{1}{m} \sum_{i=1}^m L_i$$

梯度:

$$\nabla_{\mathbf{w}} L(\mathbf{w}) = \frac{1}{m} (\mathbf{f} - \mathbf{y})^T \mathbf{X} = \frac{1}{m} (\sigma(\mathbf{X}\mathbf{w}) - \mathbf{y})^T \mathbf{X}$$

梯度推導:

為了討論如何求 $L(\mathbf{w})$ 關於 $\mathbf{w} = (w_0, w_1, \dots, w_K)^T$ 的偏導數，在此引入助記號 $z^{(i)}$ 和 $f^{(i)}$ ，公式如下。

$$z^{(i)} = \mathbf{w} \odot \mathbf{x}^{(i)} = w_1 * x_1^{(i)} + w_2 * x_2^{(i)} + \dots + w_K * x_K^{(i)} + w_0 * x_0^{(i)}$$

$$f^{(i)} = \sigma(z^{(i)})$$

$$\mathcal{L}^{(i)} = -\left(y^{(i)} \log(f^{(i)}) + (1 - y^{(i)}) \log(1 - f^{(i)})\right)$$

$$\mathcal{L}(\mathbf{w}) = \frac{1}{m} \sum_{i=1}^m \mathcal{L}^{(i)}$$

在以上公式中， $\mathcal{L}(\mathbf{w})$ 可以看成 m 個 $\mathcal{L}^{(i)}$ 的和， $\mathcal{L}^{(i)}$ 是 $f^{(i)}$ 的函數， $f^{(i)}$ 是 $z^{(i)}$ 的函數， $z^{(i)}$ 是 $\mathbf{w} = (w_1, w_2, \dots, w_K)^T$ 的函數。根據求導的四則運算法則和複合函數的連鎖律，有

$$\frac{\partial \mathcal{L}(\mathbf{w})}{\partial \mathcal{L}^{(i)}} = \frac{1}{m}$$

$$\frac{\partial \mathcal{L}^{(i)}}{\partial f^{(i)}} = -\left(\frac{y^{(i)}}{f^{(i)}} - \frac{(1 - y^{(i)})}{(1 - f^{(i)})}\right) = \frac{f^{(i)} - y^{(i)}}{f^{(i)}(1 - f^{(i)})}$$

$$\frac{\partial f^{(i)}}{\partial z^{(i)}} = \sigma(z^{(i)}) (1 - \sigma(z^{(i)})) = f^{(i)}(1 - f^{(i)})$$

$$\frac{\partial z^{(i)}}{\partial w_j} = x_j^{(i)}$$

因此，有

$$\begin{aligned} \frac{\partial \mathcal{L}(\mathbf{w})}{\partial w_j} &= \sum_{i=1}^m \frac{\partial \mathcal{L}(\mathbf{w})}{\partial \mathcal{L}^{(i)}} \times \frac{\partial \mathcal{L}^{(i)}}{\partial f^{(i)}} \times \frac{\partial f^{(i)}}{\partial z^{(i)}} \times \frac{\partial z^{(i)}}{\partial w_j} \\ &= \frac{1}{m} \sum_{i=1}^m \frac{f^{(i)} - y^{(i)}}{f^{(i)}(1 - f^{(i)})} \times f^{(i)}(1 - f^{(i)}) \times x_j^{(i)} \\ &= \frac{1}{m} \sum_{i=1}^m (f^{(i)} - y^{(i)}) x_j^{(i)} \\ &= \frac{1}{m} \sum_{i=1}^m (f_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)}) x_j^{(i)} \\ &= \frac{1}{m} \sum_{i=1}^m x_j^{(i)} (f_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)}) \end{aligned}$$

因為 $f_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)}$ 是一個數值，所以，它與向量的數乘可以交換順序即

$$(f_w(\mathbf{x}^{(i)}) - y^{(i)})x_j^{(i)} = x_j^{(i)}(f_w(\mathbf{x}^{(i)}) - y^{(i)})$$

可以看出，對於一個樣本 (\mathbf{x}, y) ， $L(\mathbf{w})$ 關於累加和 $z = \mathbf{xw}$ 的梯度（導數） $\frac{\partial L}{\partial z}$ 是 $f - y$ ，這和線性回歸的方差 $\frac{1}{2}(f - y)^2$ 關於 f 的梯度（導數）的形式是一樣的。

如果將 \mathbf{x}^i 寫成行向量的形式，那麼所有的 \mathbf{x}^i 可以按行組成一個矩陣 \mathbf{X} ，對應地，所有樣本的目標值和預測值 y^i 和 f^i 可以寫成列向量的形式，公式如下。

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^i \\ \vdots \\ \mathbf{x}^m \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^i \\ \vdots \\ y^m \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} f_w(\mathbf{x}^1) \\ f_w(\mathbf{x}^2) \\ \vdots \\ f_w(\mathbf{x}^i) \\ \vdots \\ f_w(\mathbf{x}^m) \end{bmatrix} = \begin{bmatrix} \sigma(\mathbf{x}^1 \mathbf{w}) \\ \sigma(\mathbf{x}^2 \mathbf{w}) \\ \vdots \\ \sigma(\mathbf{x}^i \mathbf{w}) \\ \vdots \\ \sigma(\mathbf{x}^m \mathbf{w}) \end{bmatrix} = \sigma(\mathbf{X}\mathbf{w})$$

將所有 $L(\mathbf{w})$ 關於 w_j 的偏導數 $\frac{\partial L(\mathbf{w})}{\partial w_j} = \frac{1}{m} \sum_{i=1}^m x_j^{(i)} (f_w(\mathbf{x}^i) - y^i)$ 寫成行向量的形式，公式如下。

$$\begin{aligned} \nabla_{\mathbf{w}} L(\mathbf{w}) &= \left[\frac{\partial L(\mathbf{w})}{\partial w_0} \quad \frac{\partial L(\mathbf{w})}{\partial w_1} \quad \frac{\partial L(\mathbf{w})}{\partial w_2} \quad \cdots \quad \frac{\partial L(\mathbf{w})}{\partial w_n} \right] \\ &= \left[\frac{1}{m} \sum_{i=1}^m x_0^{(i)} (f_w(\mathbf{x}^i) - y^i) \quad \frac{1}{m} \sum_{i=1}^m x_1^{(i)} (f_w(\mathbf{x}^i) - y^i) \quad \frac{1}{m} \sum_{i=1}^m x_2^{(i)} (f_w(\mathbf{x}^i) - y^i) \quad \cdots \quad \frac{1}{m} \sum_{i=1}^m x_n^{(i)} (f_w(\mathbf{x}^i) - y^i) \right] \\ &= \frac{1}{m} \sum_{i=1}^m [x_0^{(i)} (f_w(\mathbf{x}^i) - y^i) \quad x_1^{(i)} (f_w(\mathbf{x}^i) - y^i) \quad x_2^{(i)} (f_w(\mathbf{x}^i) - y^i) \quad \cdots \quad x_n^{(i)} (f_w(\mathbf{x}^i) - y^i)] \\ &= \frac{1}{m} \sum_{i=1}^m [x_0^{(i)} \quad x_1^{(i)} \quad x_2^{(i)} \quad \cdots \quad x_n^{(i)}] (f_w(\mathbf{x}^i) - y^i) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{m} \sum_{i=1}^m x^i (f_w(x^i) - y^i) = \frac{1}{m} \sum_{i=1}^m (f_w(x^i) - y^i) x^i \\
&= \frac{1}{m} (f_w(x^1) - y^1, f_w(x^2) - y^2, \dots, f_w(x^m) - y^m) \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^m \end{bmatrix} \\
&= \frac{1}{m} (f_w(x) - y)^T X = \frac{1}{m} (f - y)^T X
\end{aligned}$$

梯度 $\nabla_w L(w)$ 可表示為

$$\nabla_w L(w) = \frac{1}{m} (f - y)^T X = \frac{1}{m} (\sigma(Xw) - y)^T X$$

增加正則項:

$$L(w) = -\frac{1}{m} \sum_{i=1}^m (y_i \log(f_w(x_i)) + (1 - y_i) \log(1 - f_w(x_i))) + \lambda \|w\|$$

對應的梯度:

$$\nabla_w L(w) = \frac{1}{m} (f - y)^T X + 2\lambda w = \frac{1}{m} (\sigma(Xw) - y)^T X + 2\lambda w$$

決策曲線:

用 $f_w(x) = 0.5$ 來判斷，此 $xw = 0$

亦即 $w_0 + w_1 \times x_1 + w_2 \times x_2 = 0$

可以根據 w 計算一組 $\{x_1\}$ 所對應的 $\{x_2 = -w_0 / w_2 - w_1 \times x_1 / w_2\}$