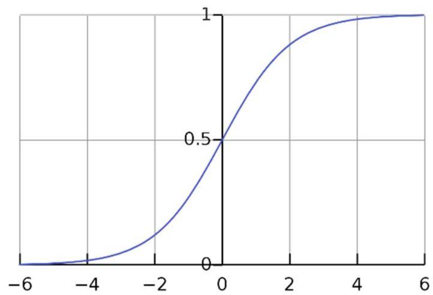
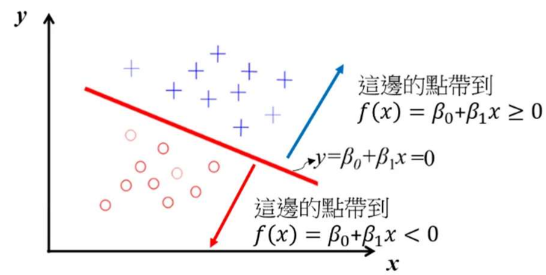
邏輯回歸:一種線性回歸的推廣，專門用於解決二分類問題。

Sigmoid 函數:$f_{\boldsymbol{w}}(\boldsymbol{x}) = \frac{1}{1+e^{-f(x)}}$，$f(x) = \boldsymbol{wx} + b$

特點:會把所有輸入值轉成 0 和 1 之間





單一樣本出現機率:

$$P(y = 1|\boldsymbol{x}) = f_{\boldsymbol{w}}(\boldsymbol{x}) = \frac{1}{1+e^{-xw}} = \sigma(\boldsymbol{xw})$$

$$P(y = 0|\boldsymbol{x}) = 1 - f_{\boldsymbol{w}}(\boldsymbol{x}) = 1 - \frac{1}{1+e^{-xw}} = 1 - \sigma(\boldsymbol{xw})$$

m 個樣本出現機率:

$$\prod_{i=1}^{m}(f_{\boldsymbol{w}}(\boldsymbol{x}_i)^{y_i})(1 - f_{\boldsymbol{w}}(\boldsymbol{x}_i))^{1-y_i})$$

代價函數:取 log

$$L(\boldsymbol{w}) = -\frac{1}{m}\sum_{i=1}^{m}(y_i \log\left(f_{\boldsymbol{w}}(x_i)\right) + (1 - y_i)\log\left(1 - f_{\boldsymbol{w}}(x_i)\right))$$

交叉熵:

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

梯度:

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

梯度推導:

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

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

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

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

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

在以上公式中，$\mathcal{L}(\boldsymbol{w})$ 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 $\boldsymbol{w} = (w_1, w_2, \cdots, w_k)^{\mathrm{T}}$ 的函數。根據求導的四則運算法則和複合函數的連鎖律，有

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

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

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

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

因此，有

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

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

$$\left(f_w(\boldsymbol{x}^{(i)}) - y^{(i)}\right)x_j^{(i)} = x_j^{(i)}\left(f_w(\boldsymbol{x}^{(i)}) - y^{(i)}\right)$$

可以看出，對於一個樣本 $(\boldsymbol{x}, y)$，$L(\boldsymbol{w})$ 關於累加和 $z = \boldsymbol{x}\boldsymbol{w}$ 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 $\boldsymbol{x}^i$ 寫成行向量的形式，那麼所有的 $\boldsymbol{x}^i$ 可以按行組成一個矩陣 $\boldsymbol{X}$，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。

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

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

$$\nabla_w L(\boldsymbol{w}) = \left[\frac{\partial L(\boldsymbol{w})}{\partial \boldsymbol{w}_0} \ \frac{\partial L(\boldsymbol{w})}{\partial \boldsymbol{w}_1} \ \frac{\partial L(\boldsymbol{w})}{\partial \boldsymbol{w}_2} \cdots \frac{\partial L(\boldsymbol{w})}{\partial \boldsymbol{w}_n}\right]$$

$$= \left[\frac{1}{m}\sum_{1=1}^m x_0^{(i)}\left(f_w(\boldsymbol{x}^i) - y^i\right) \ \frac{1}{m}\sum_{1=1}^m x_1^{(i)}\left(f_w(\boldsymbol{x}^i) - y^i\right) \ \frac{1}{m}\sum_{1=1}^m x_2^{(i)}\left(f_w(\boldsymbol{x}^i) - y^i\right) \cdots \frac{1}{m}\sum_{1=1}^m x_n^{(i)}\left(f_w(\boldsymbol{x}^i) - y^i\right)\right]$$

$$= \frac{1}{m}\sum_{1=1}^m \left[x_0^{(i)}\left(f_w(\boldsymbol{x}^i) - y^i\right) \ x_1^{(i)}\left(f_w(\boldsymbol{x}^i) - y^i\right) \ x_2^{(i)}\left(f_w(\boldsymbol{x}^i) - y^i\right) \cdots x_n^{(i)}\left(f_w(\boldsymbol{x}^i) - y^i\right)\right]$$

$$= \frac{1}{m}\sum_{1=1}^m \left[x_0^{(i)} \ \ x_1^{(i)} \ \ x_2^{(i)} \ \cdots \ x_n^{(i)}\right]\left(f_w(\boldsymbol{x}^i) - y^i\right)$$

$$= \frac{1}{m}\sum_{1=1}^{m} \boldsymbol{x}^i\left(f_{\boldsymbol{w}}(\boldsymbol{x}^i) - y^i\right) = \frac{1}{m}\sum_{1=1}^{m}\left(f_{\boldsymbol{w}}(\boldsymbol{x}^i) - y^i\right)\boldsymbol{x}^i$$

$$= \frac{1}{m}\left(f_{\boldsymbol{w}}(\boldsymbol{x}^1) - y^1, f_{\boldsymbol{w}}(\boldsymbol{x}^2) - y^2, \cdots, f_{\boldsymbol{w}}(\boldsymbol{x}^m) - y^m\right)\begin{bmatrix}\boldsymbol{x}^1\\ \boldsymbol{x}^2\\ \vdots\\ \boldsymbol{x}^m\end{bmatrix}$$

$$= \frac{1}{m}(\boldsymbol{f}_{\boldsymbol{w}}(\boldsymbol{x}) - \boldsymbol{y})^{\mathrm{T}}\boldsymbol{X} = \frac{1}{m}(\boldsymbol{f} - \boldsymbol{y})^{\mathrm{T}}\boldsymbol{X}$$

梯度 $\nabla_{\boldsymbol{w}}L(\boldsymbol{w})$ 可表示為

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

增加正則項:

$$L(\boldsymbol{w}) = -\frac{1}{m}\sum_{i=1}^{m}(y_i\log\left(f_{\boldsymbol{w}}(x_i)\right) + (1-y_i)\log\left(1 - f_{\boldsymbol{w}}(x_i)\right) + \lambda\|\boldsymbol{w}\|^2$$

對應的梯度:

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

決策曲線:

用$f_w(\boldsymbol{x}) == 0.5$來判斷，此$\boldsymbol{x}\boldsymbol{w} = 0$

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

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