玉山金控 E.SUN FHC

Logistic Regression

劉柏均

2018/05/22

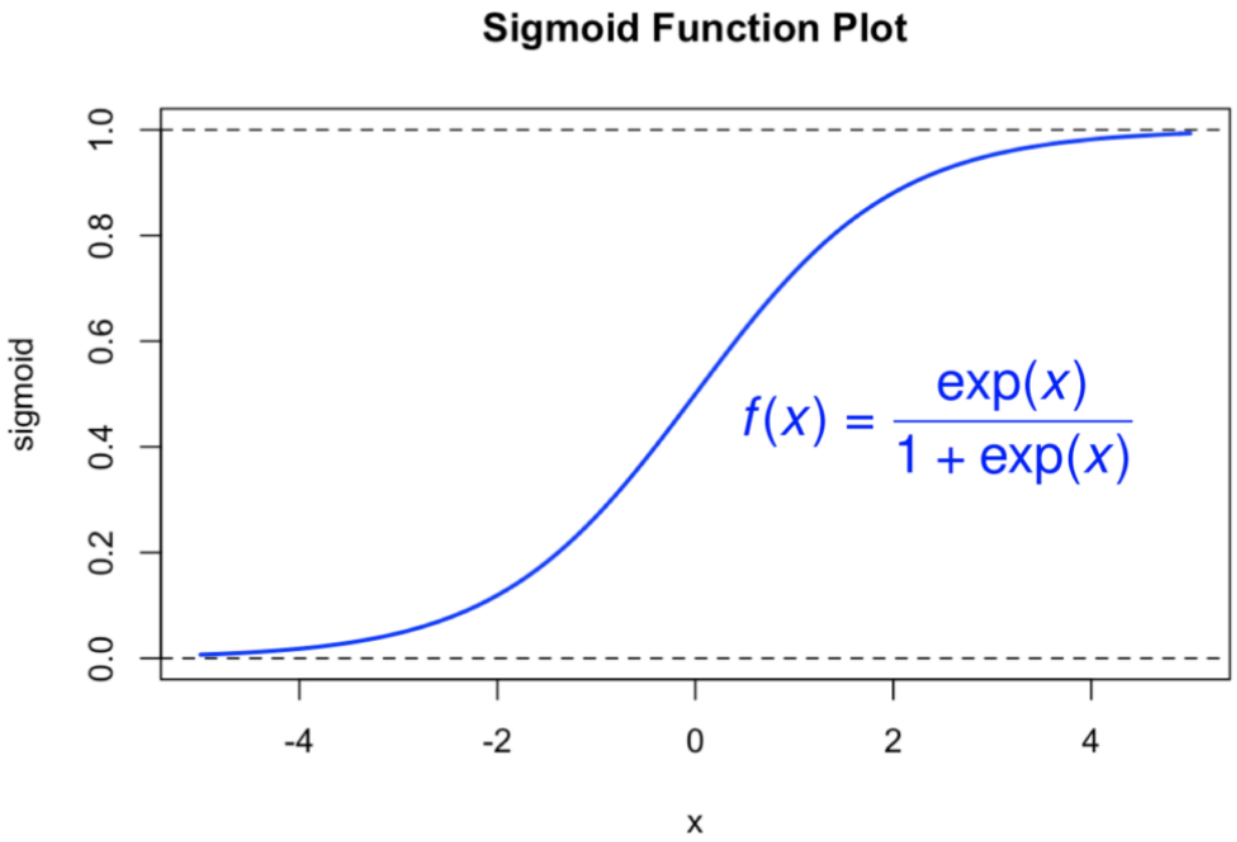
Logistic Regression

- 本質：分類模型
- 用途：用於二元分類或者多元分類
- 精神：當X變動一單位時，對於勝算比的影響

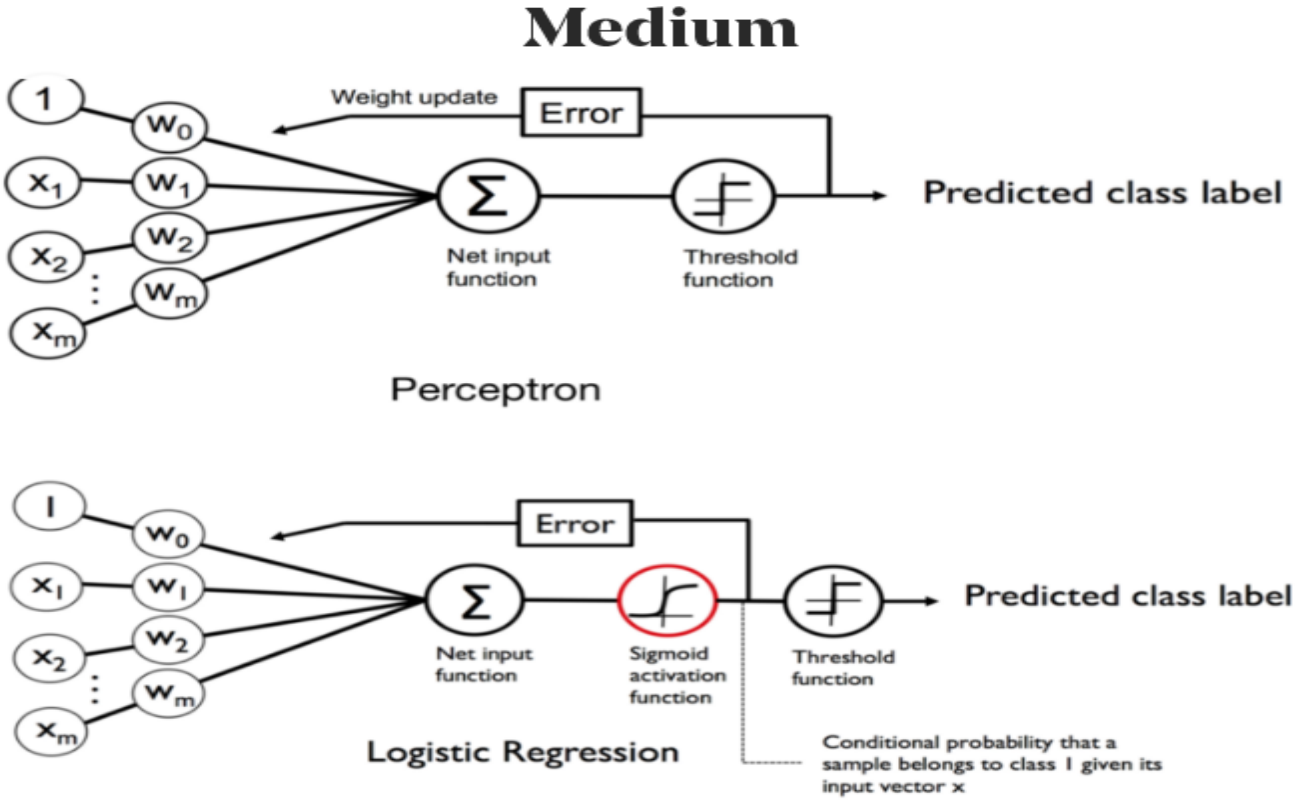
$$\log\left(\frac{P(Y=1|X)}{1-P(Y=1|X)}\right) = \beta_0 + \beta_1 X$$

Log Odds (對數勝率)　　線性迴歸

Logistic Regression



Logistic Regression



Perceptron以及Logistic Regression模型的差異

Logistic Regression

運用「最大概似法」估計參數

由於 P(*Y* = 1|*X*) 未知，無法運用最小平方法估計。

	X	*Y*
個體 1	X_1	Y_1
個體 2	X_2	Y_2
個體 3	X_3	Y_3
個體 4	X_4	Y_4
…	…	…
個體 *n*	X_n	Y_n

$$p_1 \equiv P(Y=1|X_1) = \frac{\exp(\beta_0+\beta_1 X_1)}{1+\exp(\beta_0+\beta_1 X_1)}$$

若 $Y_1 = 1$，則得到 p_1 分數；
若 $Y_1 = 0$，則得到 $1-p_1$ 分數。
該個體獲得分數 = $(p_1)^{y_1}(1-p_1)^{1-y_1}$

$$p_n \equiv P(Y=1|X_n) = \frac{\exp(\beta_0+\beta_1 X_n)}{1+\exp(\beta_0+\beta_1 X_n)}$$

找到一組參數 $(\hat{\beta}_0, \hat{\beta}_1)$ 最大化概似函數 $\mathcal{L}(\beta_0,\beta_1) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}$ 。

分數乘積

Logistic Regression

- DEMO

Logistic Regression

廣義邏輯斯迴歸模型(generalized logistic regression model)

此模型首先指定某一組為參考組，接著其他組一一與此參考組做比較，其數學式如下：

$$\log(\frac{\pi_j}{\pi_1}) = \alpha_j + \beta_j x, j = 2,...,J$$

若反應變數分三類，例如不重要、中等重要、很重要，則可得兩個數學式如下：

$$\log(\frac{\pi_{中等重要}}{\pi_{不重要}}) = \alpha_2 + \beta_2 x, 及 \log(\frac{\pi_{很重要}}{\pi_{不重要}}) = \alpha_3 + \beta_3 x,$$

Thank you