

## Logistic Regression

### 一、模型：

特性：依變數(Y) 為二分類的反應數，以 1 及 0 代表。

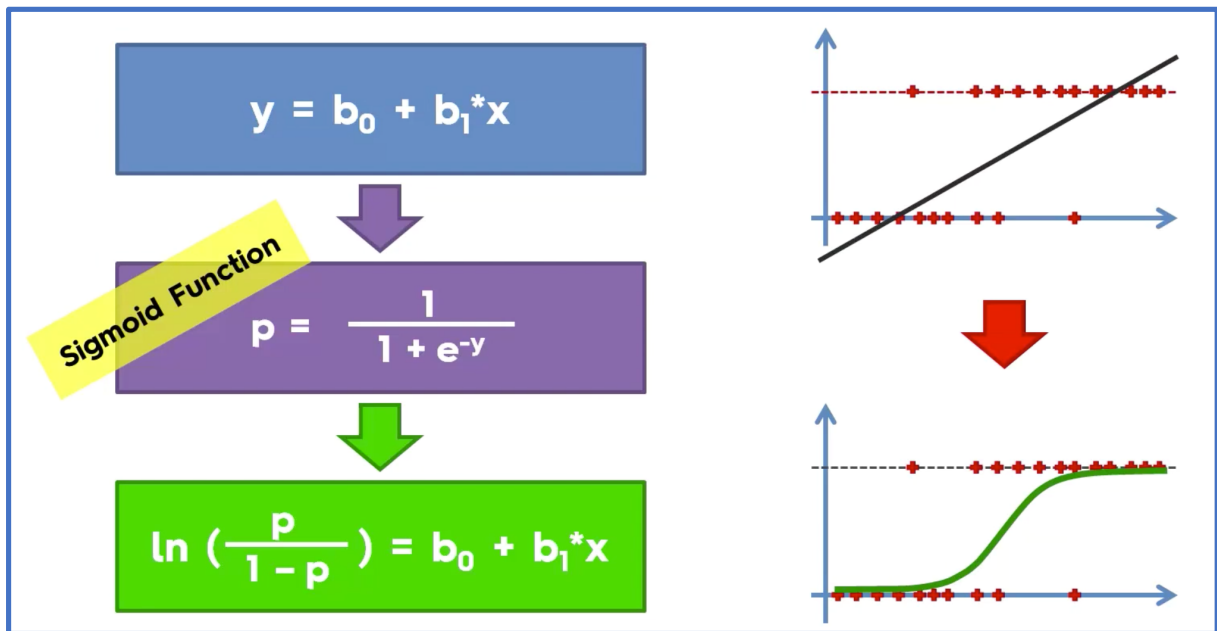
假設  $y = \ln(odds) = \beta_0 + \beta_1 X$ ,  $odds = \text{勝算} = \frac{\text{prob}(y=1)}{\text{prob}(y=0)}$

Model :

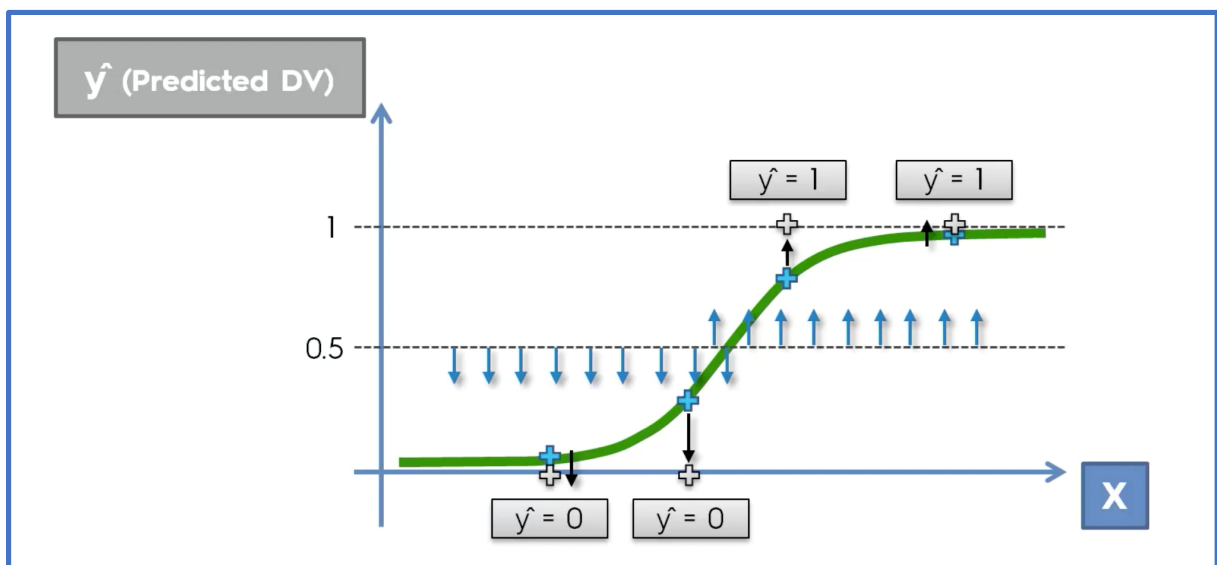
$$Y_i = E\{Y_i\} + \varepsilon_i$$

$$E\{Y_i\} = p_i = \frac{1}{1+e^{-y}}, y = \beta_0 + \beta_1 X_i, i = 1, \dots, n$$

說明:注意 Sigmoid Function 算出預測機率



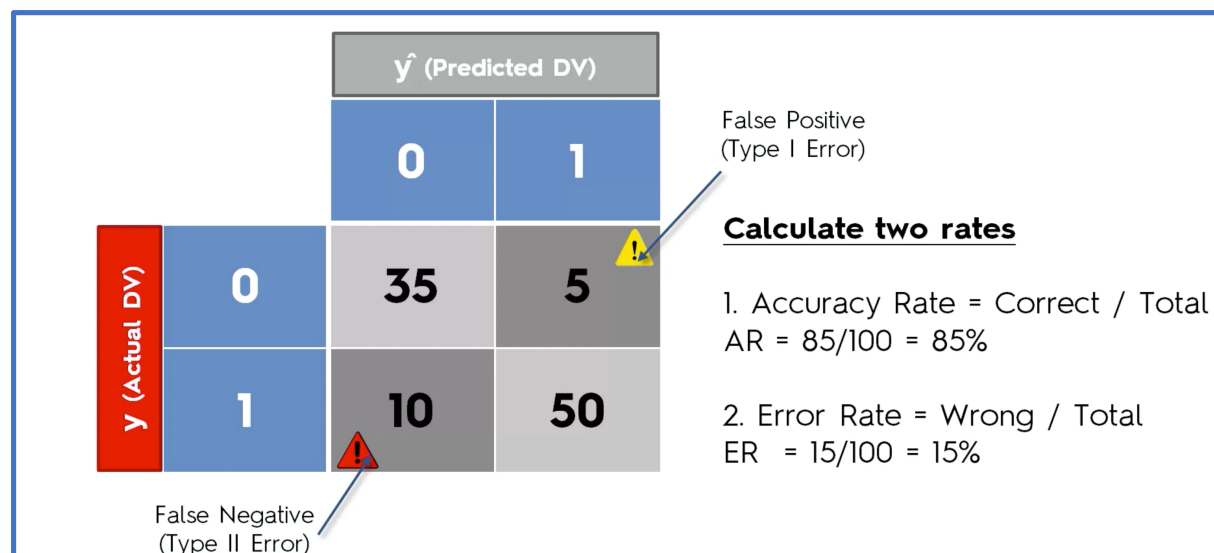
說明:用 0.5 作為切分點，將資料分群



### 二、Confusion Matrix

說明:

混淆矩陣是對有監督學習分類算法準確率進行評估的工具。通過將模型預測的數據與測試數據進行對比，使用準確率，覆蓋率和命中率等指標對模型的分類效果進行度量。



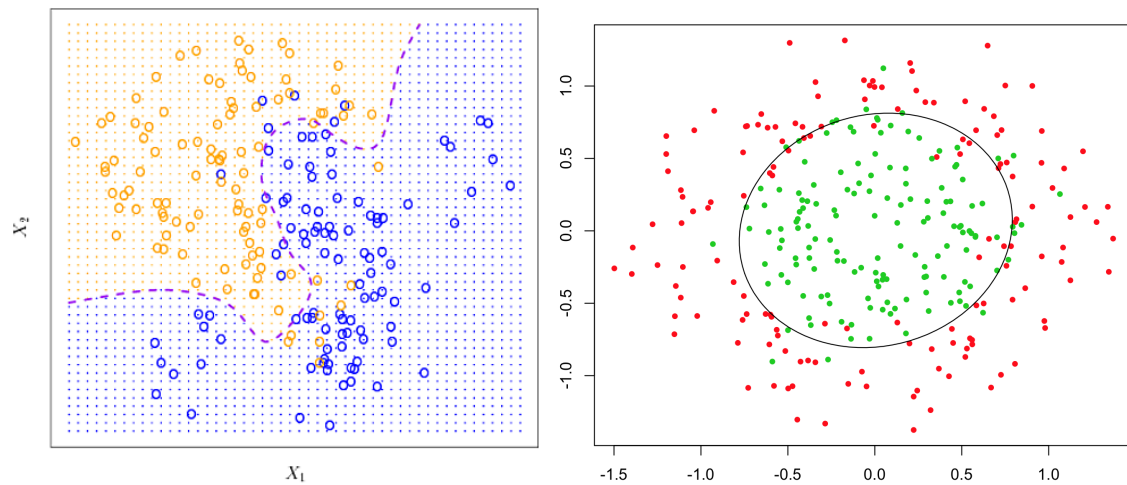
假設檢定表格複習

		根據研究結果的判斷	
		拒絕 $H_0$ (實際上拒絕零假設)	接受 $H_0$ (實際上接受零假設)
真實情況	$H_0$ 是真實的 (理論上應接受零假設)	錯誤判斷 (陽性判斷錯誤 偽陽性、 <b>type-1 error</b> )	正確判斷
	$H_0$ 是錯誤的 (理論上不接受零假設)	正確判斷	錯誤判斷 (陰性判斷錯誤 偽陰性、 <b>type-2 error</b> )

### 三、Decision Boundary

名詞解釋:

在進行決策的過程中，常將特徵空間劃分成若干空間，每兩個決策空間的交界處即為決策邊界。如直線或曲線即為二維空間的決策邊界，平面或曲面即為三維空間的決策邊界。



說明:下圖解釋如何畫出 logistic regression 的 Decision Boundary

