Logistic Regression

- Lets now focus on the binary classification problem in which ○ y can take on only two values, 0 and 1.
    - ○ x is a vector of real-valued features, < x1 ... xn >
- We could approach the classification problem ignoring the fact that y is discrete-valued, and use our old linear regression algorithm to try to predict y given x.
    - ○ However, it doesn't make sense for f(x) to possibly take values larger than 1 or smaller than 0 when we know that y ∈ {0, 1}.
- Since the output must be 0 or 1, we cannot directly use a linear model to estimate f(x).
- Furthermore, we would like to f(x) to represent the probability P(C1|x). Lets call it p.
- We will model the log of the odds of the probability p as a linear function of the input x.

$$odds = \frac{p}{1-p}$$

ln (odds of p) = ln (p/(1-p)) = **w**.x

If there is a 75% chance that it will rain tomorrow, then the odds of it raining tomorrow are 3 to 1. (¾)/¼=3/1.

- This is the logit function. I.e. logit(p) = ln (p/(1-p))
- By applying the *inverse of the logit function, that is* the logistic function, on both sides, we get:

$$\text{logit}^{-1}\left(\ln(p/(1-p))\right) = \text{sigmoid}\left(\ln(p/(1-p))\right) = p$$

- Applying it on the RHS as well, we get

$$p = \text{logit}^{-1}(\mathbf{w}.\mathbf{x}) = 1 / (1 + e^{-\mathbf{w}.\mathbf{x}})$$

Odds & Odds Ratios

The odds has a range of 0 to ∞ with values :

- greater than 1 associated with an event being more likely to occur than not to occur and

- values less than 1 associated with an event that is less likely to occur than not occur.

$$\ln(odds) = \ln\left(\frac{p}{1-p}\right) = \ln(p) - \ln(1-p)$$

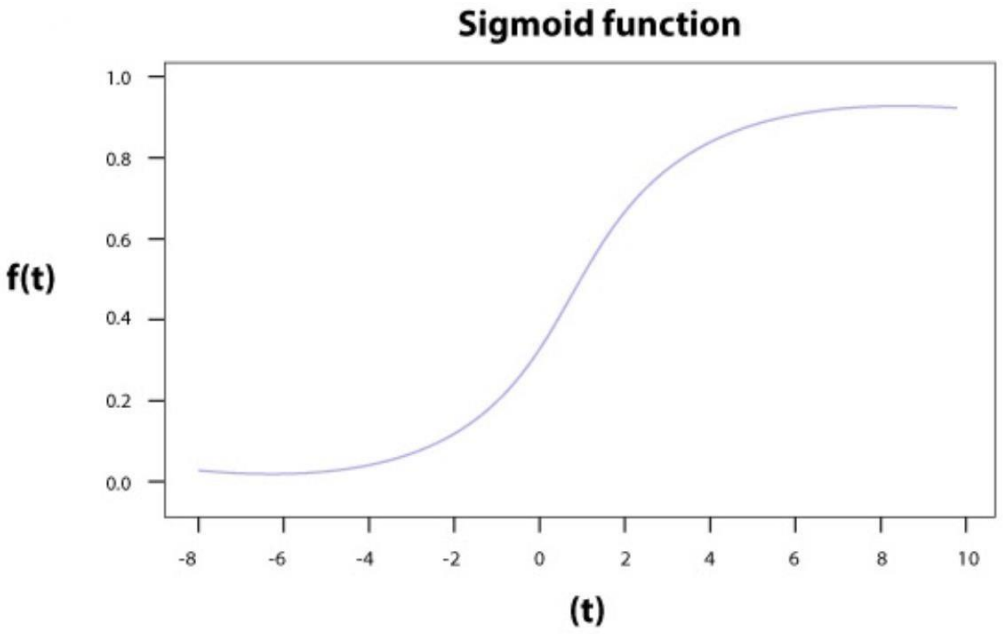
- The logit is defined as the log of the odds(-∞to+∞)

  As β.x gets really big, p approaches 1
  As β.x gets really small, p approaches 0

What is the Sigmoid Function?

It is a mathematical function having a characteristic that can take any real value and map it to between 0 to 1 shaped like the letter “S”. The sigmoid function also called a logistic function.

$$Y = 1 / 1+e^{-z}$$



So, if the value of z goes to positive infinity then the predicted value of y will become 1 and if it goes to negative infinity then the predicted value of y will become 0. And if the outcome of the sigmoid function is more than 0.5 then we classify that label as class 1 or positive class and if it is less than 0.5 than we can classify it to negative class or label as class 0.