Logistic Regression

What is logistic Regression?

⇒ Logistic Regression is a supervised learning classification algorithm used to predict the probability of a target variable.

⇒ The nature of target or dependent variable is dichotomus, which means there would be only two possible class.

⇒ Mathematically, a Logistic regression model predicts $P(Y=1)$ as a function of X.

Types of logistic Regression:

Binary or Binomial
(1/0, Yes/No)

Multinomail
[Unordered types]
(TypeA, TypeB, TypeC)

Ordinal
[Ordered types]
(poor, good, very good, excellent)

Sigmoid Activation:

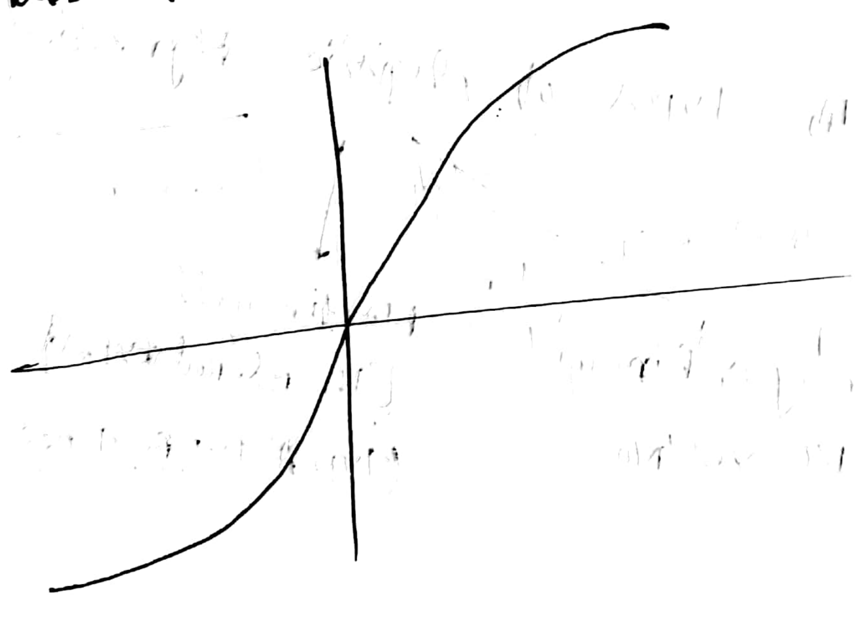
In order to map predicted values to probabilities, we use the sigmoid function.

Math: $y = F(x) = \frac{1}{(1+e^{-x})}$

$F(x)$ = Output between 0 and 1 (probability estimate)

x = input the function (your Algorithm's prediction : e.g: $mx+b$)

e = base of natural log.



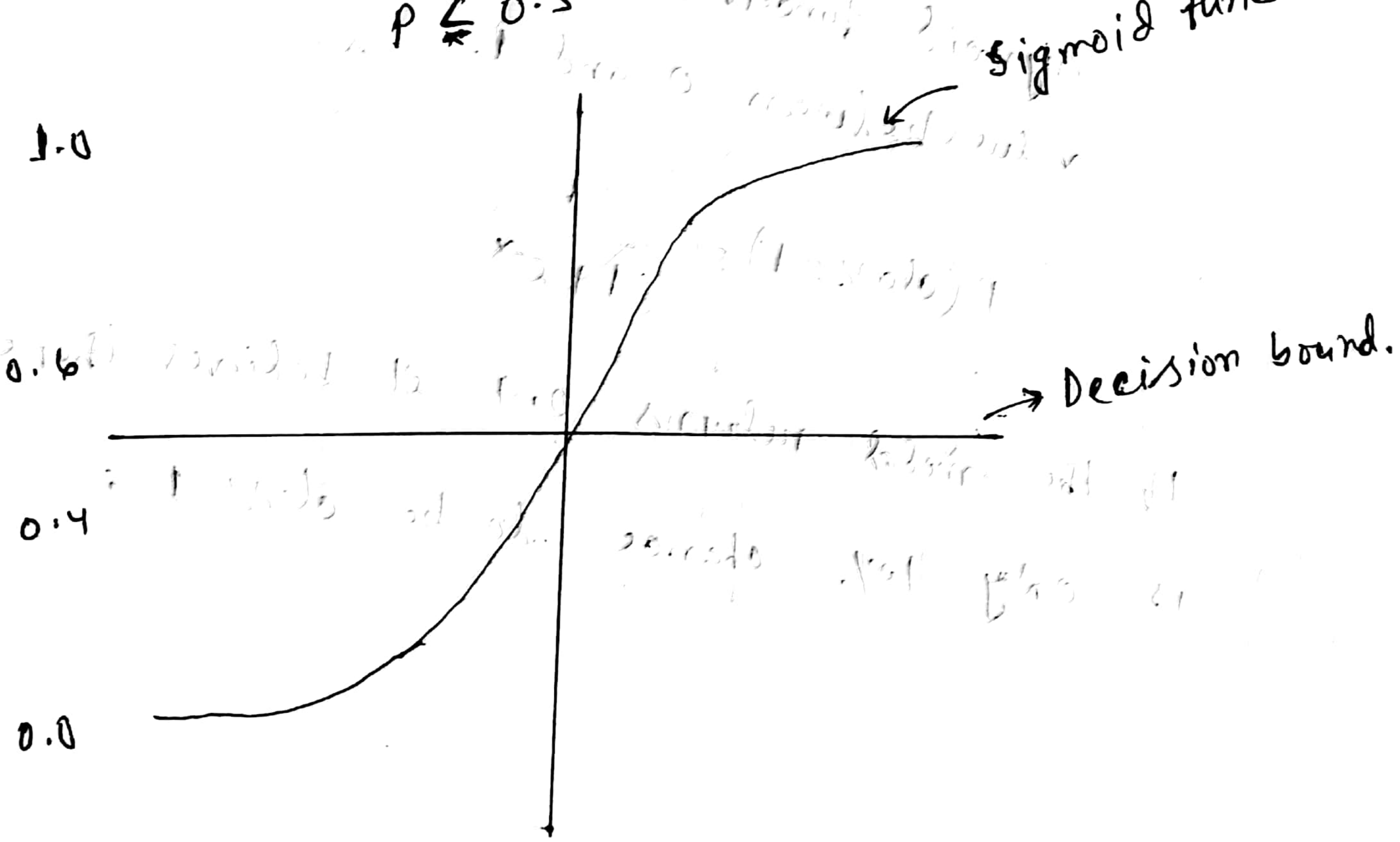
Decission boundary:

Our prediction function return a probability score between 0 and 1.

⇒ We have to select a thresold value or tripping point above which we will classify values into class 1 and below which we classify value into class 2!

$p \geq 0.5$; class = 1 [threshold = 0.5]

$p \leq 0.5$ class = 0



Making Predictions:

Using our knowledge of sigmoid functions and decision boundaries, we can now write a prediction function.

Math: Using the same Multiple Linear Regression

$$x = w_0 + w_1 x_1 + w_2 x_2 \cdots + w_n x_n$$

We will transform the output using the sigmoid function to return a probability value between 0 and 1.

$$P(class = 1) = \frac{1}{1 + e^{-x}}$$

If the model returns 0.4 it believes there is only 40% chance to be class 1.

Cost function:

Instead of Mean Squared Error, we use a cost function called "Cross Entropy". Also known as log loss.

Math:

We used before in linear Regression:

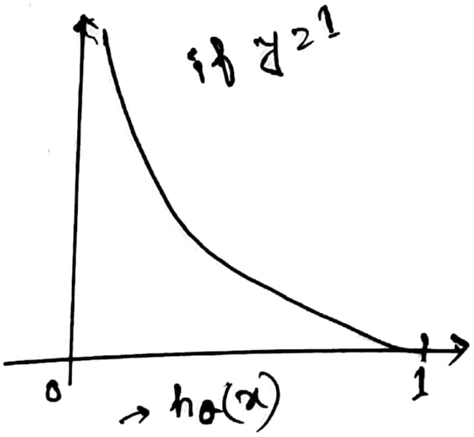
$$J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x_i) - y_i)^2 \; ; \theta_0 = 0$$

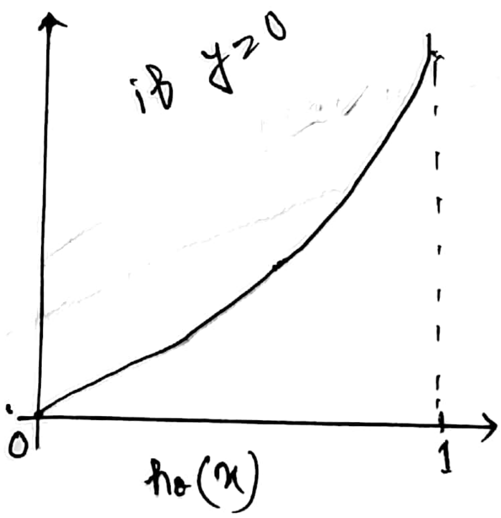
Now,

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} cost(h_\theta(x_i), y_i)$$

$$cost(h_\theta(x), y) = -\log(h_\theta(x)) \quad \text{if } y = 1$$

$$cost(h_\theta(x, y)) = -\log(1 - h_\theta(x)) \quad \text{if } y = 0$$





These smooth ~~form~~ monotonic functions (always increasing or always decreasing) make it easy to calculate the minimized cost.

Above functions compressed into One:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y_i \log(h_\theta(x_i)) + (1-y_i)\log(1-h_\theta(x_i))\right]$$

if $Y=0$ first side cancels out, if $y=1$, second side cancels out.