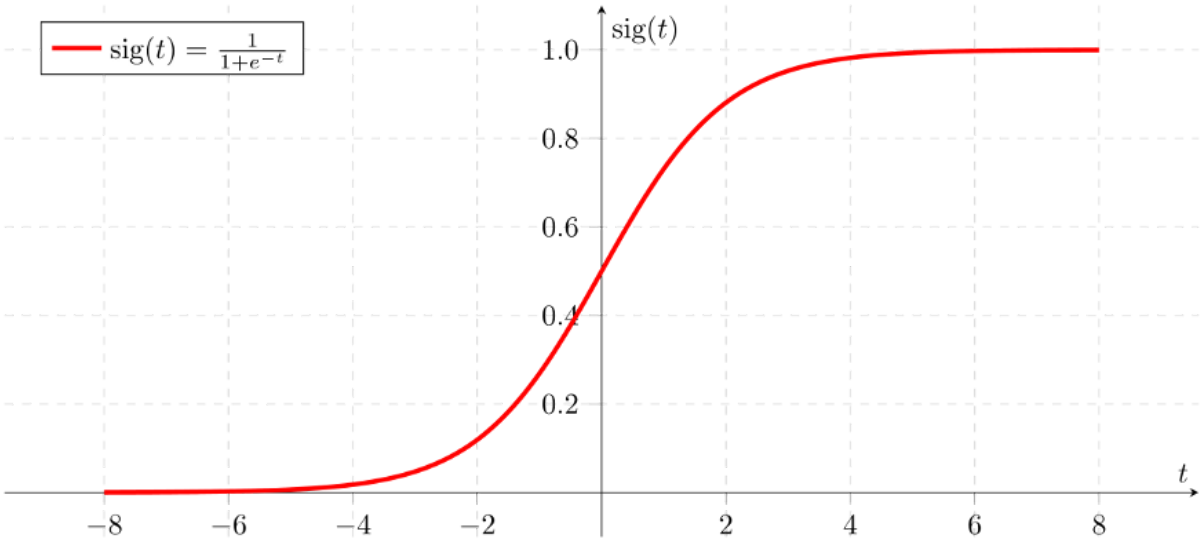
Logistic Regression

- Logistic Regression is the appropriate regression analysis to conduct when the dependent variable is binary.
- In a lot of ways, linear regression and logistic regression are similar. But, the biggest difference lies in what they are used for. Linear regression algorithms are used to predict/forecast values but logistic regression is used for classification tasks.
- Examples of Logistic Regression are classifying whether an email is spam or not, classifying whether a tumour is malignant or benign, classifying whether a website is fraudulent or not, etc.

Sigmoid Function (Logistic Function)



$$z = \theta_0 + \theta_1 \cdot x_1 + \theta \cdot x_2 + \cdots$$

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

$$h = g(z) = \frac{1}{1 + e^{-z}}$$

If 'Z' goes to infinity, h(predicted) will become 1 and if 'Z' goes to negative infinity, h(predicted) will become 0.

Decision Boundary

The decision Boundary for Logistic Regression is considered at y=0.5.
Say, if predicted_value ≥ 0.5, then classify email as spam else as not spam.

Cost Function

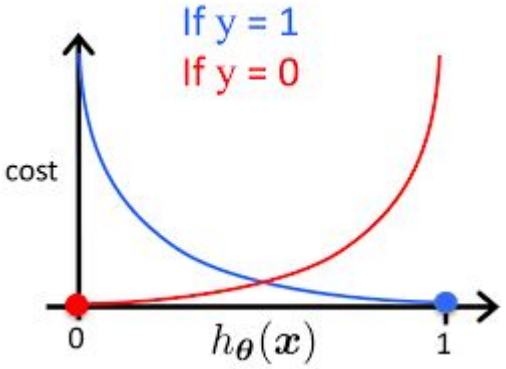
The cost function for linear regression is as follows:

$$\text{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

Simplified form of Cost function is:

$$J(\theta) = -\frac{1}{m}\sum\left[y^{(i)}\log(h\theta(x(i))) + \left(1 - y^{(i)}\right)\log(1 - h\theta(x(i)))\right]$$

Where h is called a hypothesis.



Gradient Descent

The main goal of Gradient descent is to **minimize the cost value.** i.e. min J(***θ***).
Now to minimize our cost function we need to run the gradient descent function on each parameter i.e.

$$\theta j := \theta j - \alpha \frac{\partial}{\partial \theta j} J(\theta)$$

Want $\min_\theta J(\theta)$:

Repeat {

$$\theta_j := \theta_j - \alpha \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

(simultaneously update all θ_j)

}

Image :Andrew Ng Course

.