Reference : <https://towardsdatascience.com/a-study-of-classification-problems-using-logistic-regression-and-an-insight-to-the-admissions-ec69ddf93f36>

Supervised Learning is a machine learning technique in which we associate our inputs with our targets in the given dataset. We already have a definite intuition regarding our final output. We broadly have two types of Supervised Learning problems, “Regression” and “Classification”. We will be discussing about Classification problems in this article. A Classification problem is a problem in which we separate our input data into discrete categories. By discrete, we mean separate classes.

The algorithm which generally resolves a dataset into discrete categories is defined as a classifier

### **Activation Function and Mapping Function involved:**

First of all, consider the function that maps our output and input and is hence used for computing predictions while calculating the Squared Error Function. Let us call that function a Mapping Function or a Hypothesis Function, **h(X)**. The term “Hypothesis” was coined for various historical reasons. Also, the term “Logistic” is a synonym for Sigmoid. Thus, we use a Sigmoid activation function with an “S Shaped curve” in this algorithm.

Mathematically, we define our function as: S(t) = 1 / ( 1 + e^-t)

The Sigmoid function also has several other mathematical properties. For example, we can represent the derivative of the function by the function itself. S’(t) = S(t)(1-S(t))

This, the Sigmoid/Logistic function simplifies the mathematics involved during optimization and is an ideal choice for a small scale classification problem. Using the Sigmoid function, we can effectively limit our range so that it represents probability effectively. We can represent our Hypothesis/Mapping function mathematically as:

h(x) = ThetaTx and z = ThetaTx and g(z) = 1 / ( 1 + e^-z)

This is essentially a very basic linear cost function, just that it’s coupled with a sigmoid function.

Our Hypothesis function gives us the numerical value for the probability any given event happening. We can define the range of the mapping function as: **0≤h(X)≤1**.

### **Cost Function Intuition:**

A cost function is a mathematical function to estimate how well the hypothesis suits the input set and target set. Graphically, we know that a better cost function will fit our given data better. It is generally denoted as a function of our parameters. In this case, we refer to it as **J(θ)**.   
In this algorithm, we use a **logarithmic cost function**which is derived from the principle of Maximum Likelihood Estimation(M.S.E) to ensure that the function we get as an output is a **Convex function.**

[1 0] [0.4]

[1 1] [50]