COURSE NAME: ARTIFICIAL INTELLIGENCE
COURSE CODE: CIS 412

SYED TANGIM PASHA

LECTURER,

DEPARTMENT OF COMPUTING AND INFORMATION SYSTEM (CIS)

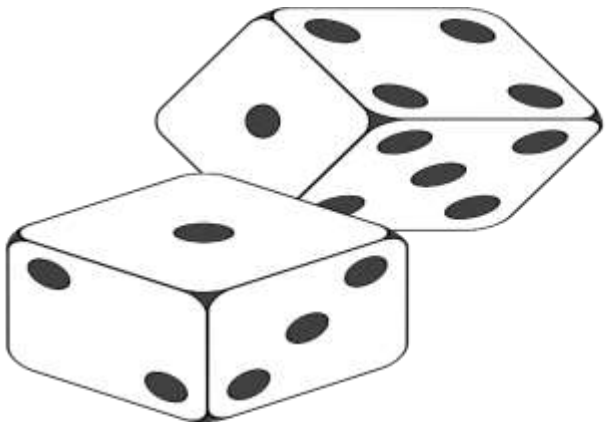
DAFFODIL INTERNATIONAL UNIVERSITY (DIU)

DHAKA,BANGLADESH

CONTINUOUS vs DISCRETE DATA

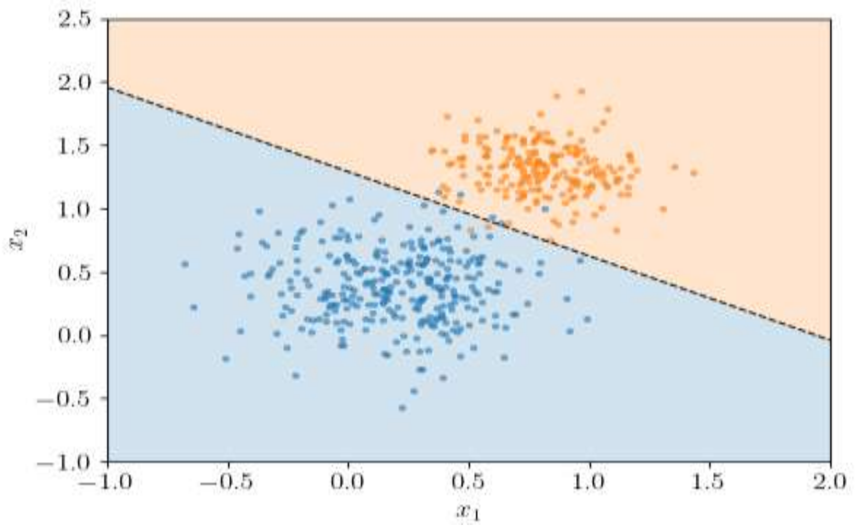
- **Continuous:** Continuous data is data that falls in a constant sequence (range). Ex: Person's height, length of a leaf etc.
- Continuous data is measured.

- **Discrete:** Discrete data can only take certain values.
- Discrete data is counted.
- Only has the values 2,3,4,5,6,7,8,9,10,11 and 12.



LOGISTIC REGRESSION

- **Logistic Regression:** Logistic Regression is a supervised learning classification algorithm used to predict the probability of a target variable.
- It is a classification algorithm.
- Works on categorical or discrete value.
- It gives the probabilistic values which lie between 0 and 1.

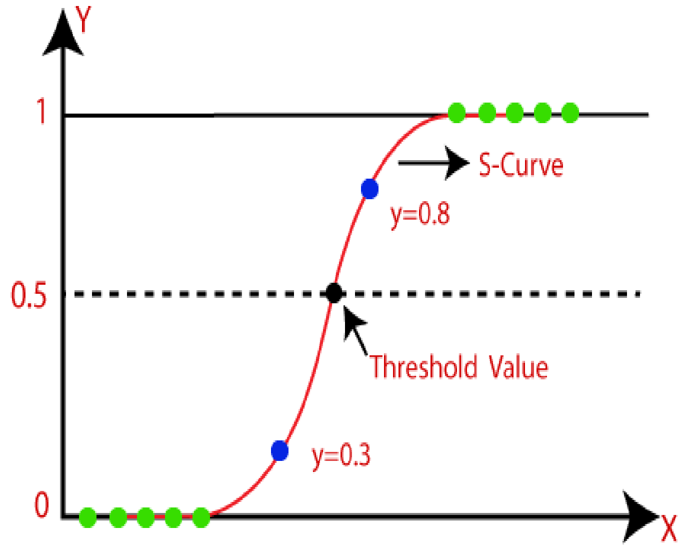


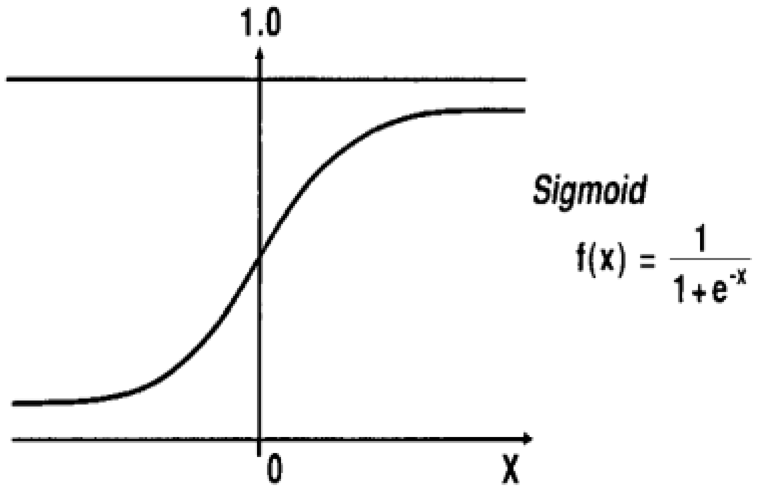
LOGISTIC REGRESSION (TYPES)

- **Binomial:** In binomial logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, Yes or No etc.
- **Multinomial:** In multinomial logistic regression, there can be 3 or more possible unordered types of the dependent variable such as ‘cat’, ‘dog’, or ‘sheep’.
- **Ordinal:** In ordinal logistic regression, there can be 3 or more possible ordered types of dependent variables such as ‘Low’, ‘Medium’ or ‘High’.

SIGMOID FUNCTION

- Sigmoid Function: The sigmoid function is a mathematical function used to map the predicated values to probabilities.
- It maps any real value into another value within a range of 0 and 1.
- The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit. So, it forms a curve like 'S' form. The S-form curve is called the Sigmoid Function or the logistic function.
- In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.





HYPOTHESIS FUNCTION

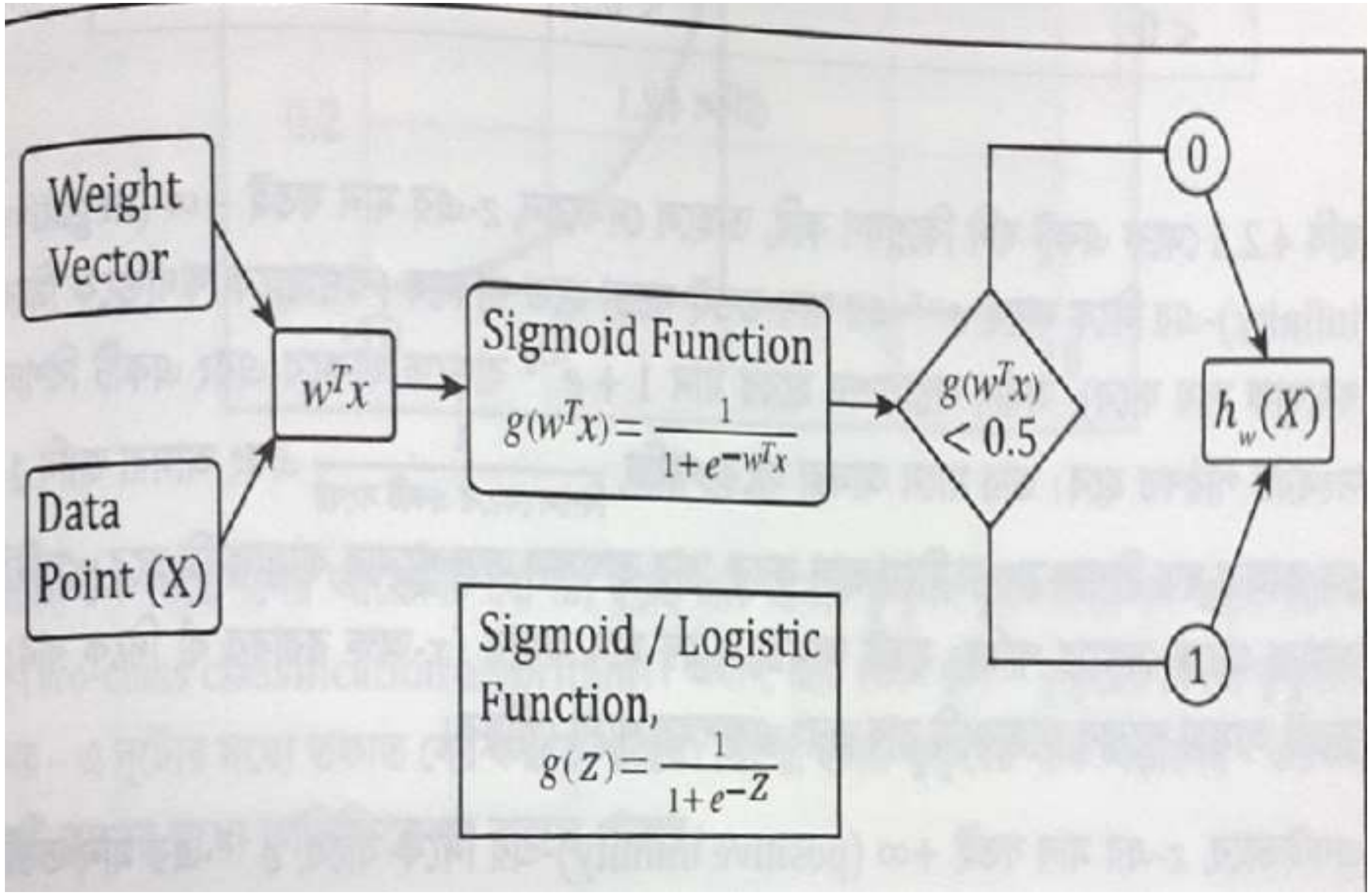
$$h_\theta(x) = \theta^T x$$

$$= g(\theta^T x)$$

$$= g(z)$$

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

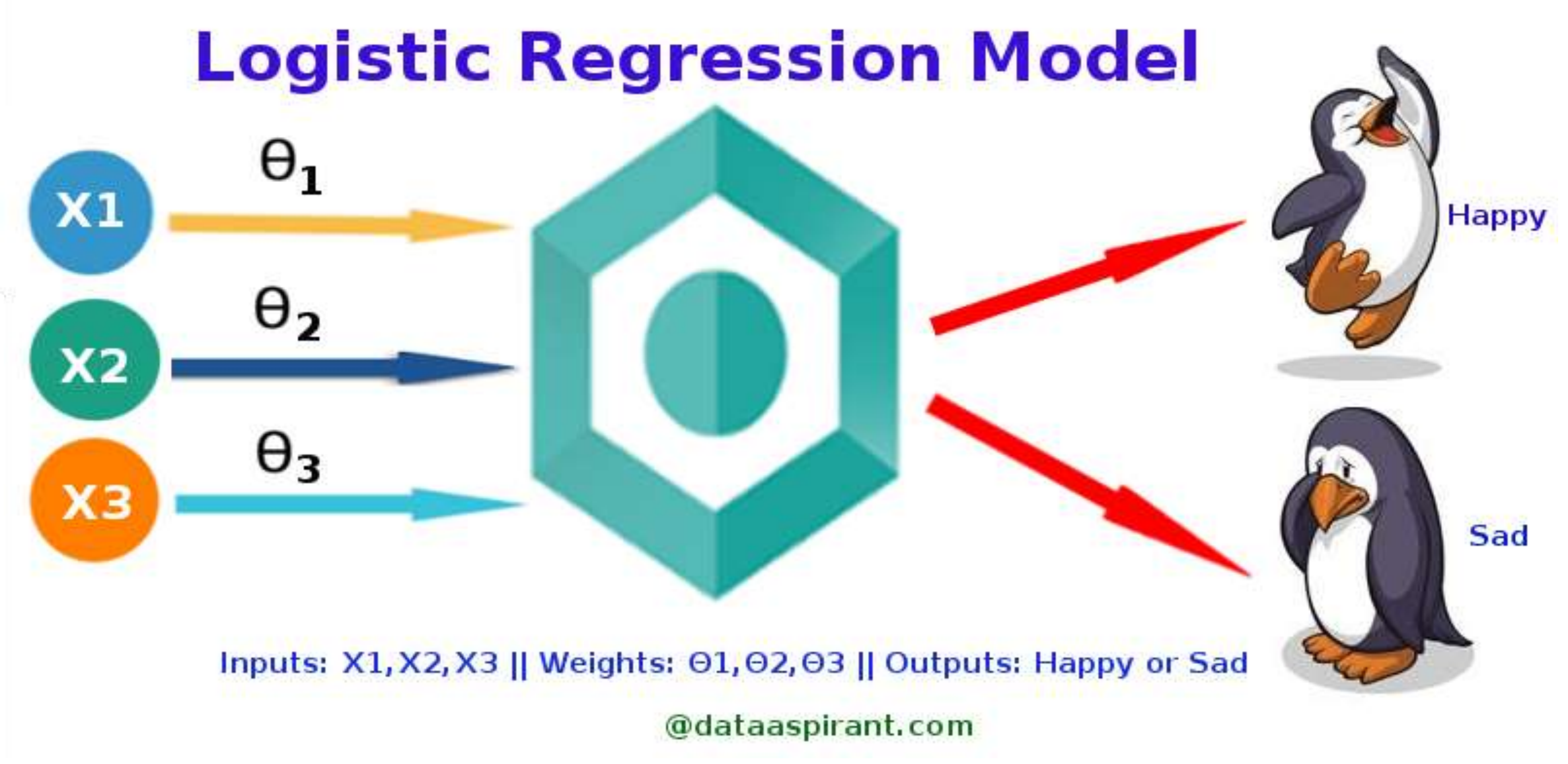
HYPOTHESIS FUNCTION



HYPOTHESIS FUNCTION

$w^T x$	$g(w^T x) = \frac{1}{1 + e^{-w^T x}}$ (এটি হাইপোথিসিস থেকে প্রাপ্ত মান)	$h_w(x)$ (এটি লজিস্টিক রিগ্রেশনের ফাইনাল আউটপুট)
0	0.5	1
> 0	> 0.5	1
< 0	< 0.5	0

LOGISTIC REGRESSION



WHY CALLED IT LOGISTIC REGRESSION ??

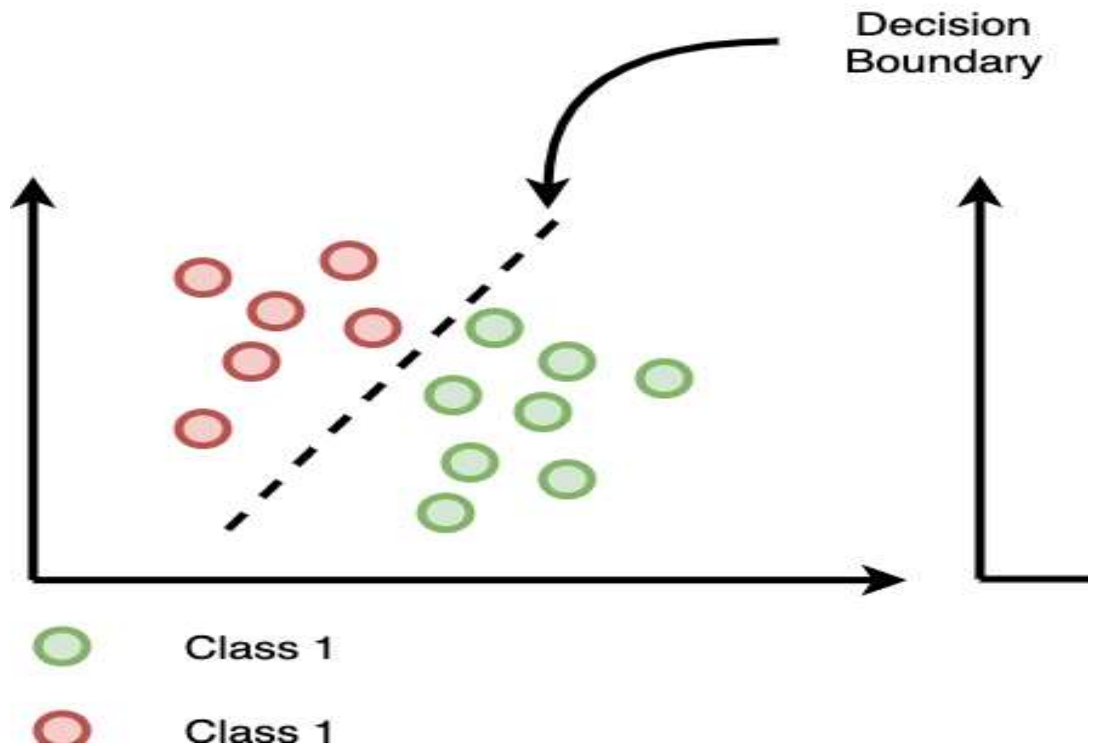
- Why we called it logistic regression?? Whereas it is a classification algorithm!!

- **Answer:** As the hypothesis function of Logistic Regression gives us the continuous value like Linear Regression, but final hypothesis value is categorical and that's why it is called as Logistic Regression but it is a classification algorithm.

As Sigmoid function is non-linear, then how Logistic Regression works as a Linear Classifier?

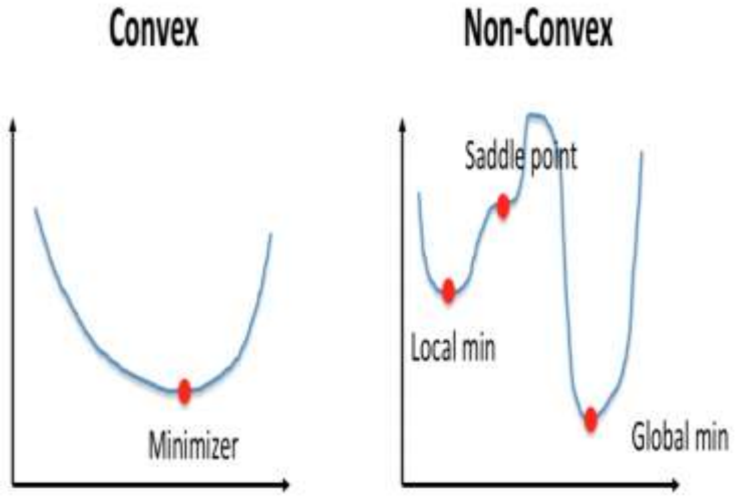
DECISION BOUNDARY

- **Decision Boundary:** A decision boundary is a line, where all samples of one class are on one side of that line, and all samples of the other class are on the opposite side of the line. The line separates one class from the other.



CONVEX vs NON-CONVEX

- **Convex & Non-Convex**: Convex and nonconvex are optimization problems. The basic difference between the two categories is that:
- **convex optimization** there can be only one optimal solution, which is globally optimal or you might prove that there is no feasible solution to the problem.
- While in **nonconvex optimization** may have multiple locally optimal points and it can take a lot of time to identify whether the problem has no solution or if the solution is global.



COST FUNCTION

$$Cost\{h_w(x^{(i)}), y^{(i)}\} = -y^{(i)} \log(h_w(x)) - (1 - y^{(i)}) \log(1 - h_w(x))$$

$$Cost\{h_w(x^{(i)}), y^{(i)}\} = \begin{cases} -\log(h_w(x)), & if\ y = 1 \\ -\log(1 - h_w(x)), & if\ y = 0 \end{cases}$$

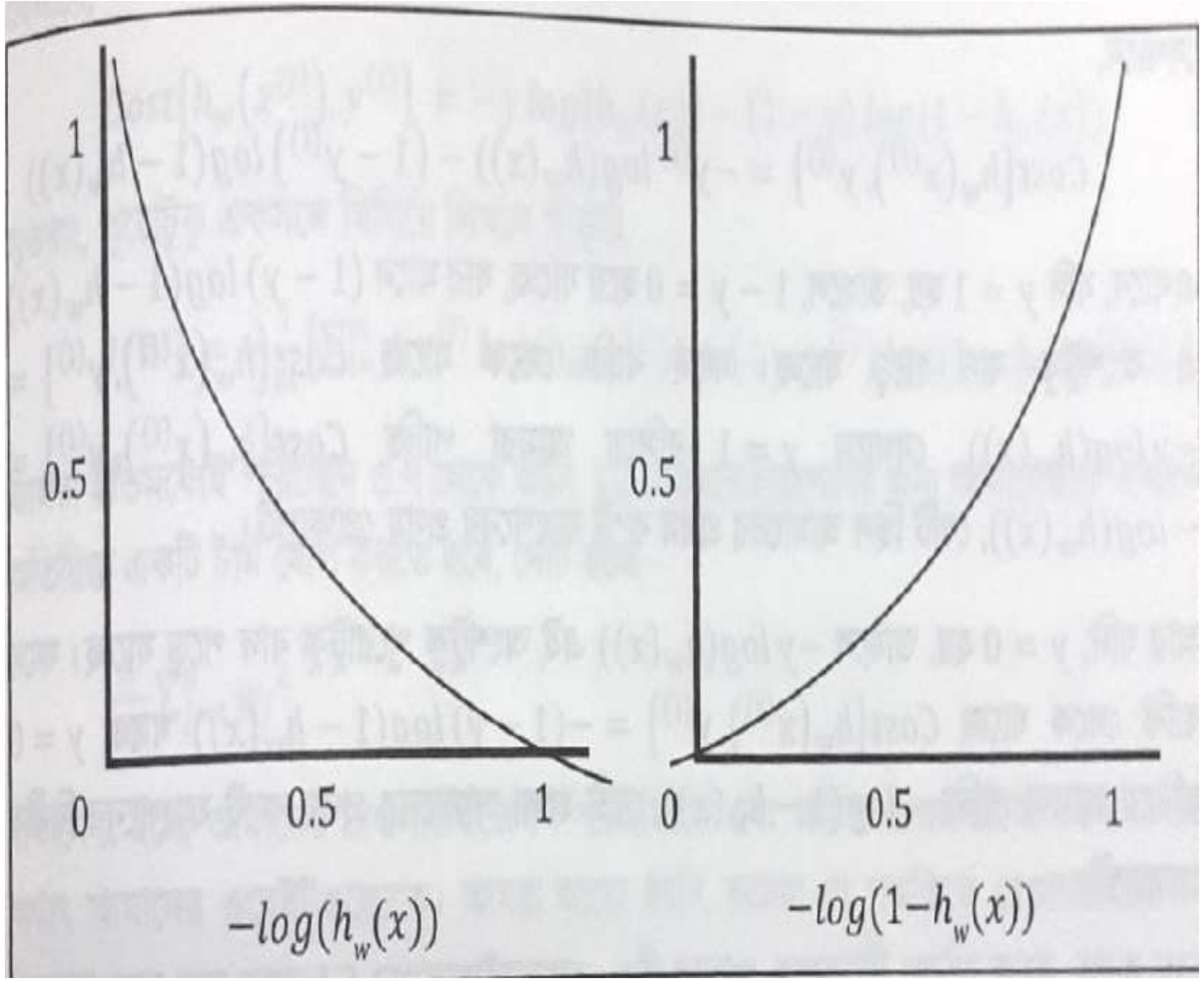
$$\begin{aligned} J(\theta) &= \frac{1}{m} \sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)}) \\ &= -\frac{1}{m} [\sum_{i=1}^{m} y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))] \end{aligned}$$

COST FUNCTION & GRADIENT DESCENT

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$
$$= -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))]$$

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

COST FUNCTION



Y	$h_w(x)$	Cost
0	0	0
	$\cong 1$	$\cong \infty$
1	1	0
	$\cong 0$	$\cong \infty$