4: Logistic Regression

Linear vs. Logistic Regression

- Linear regression is well suited when the response variable y is continuous.
- What about when y is discrete ?
  Example : y is binary (i.e) $y \in \{0, 1\}$
  This is precisely handled by Logistic Regression.

Examples :

- Image / video : Face detection
  (Person present / absent)
- Medical imaging : X-ray, CT scan, MRI scan
  (Disease present / absent)

Logistic function

The Logistic function is given below.

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

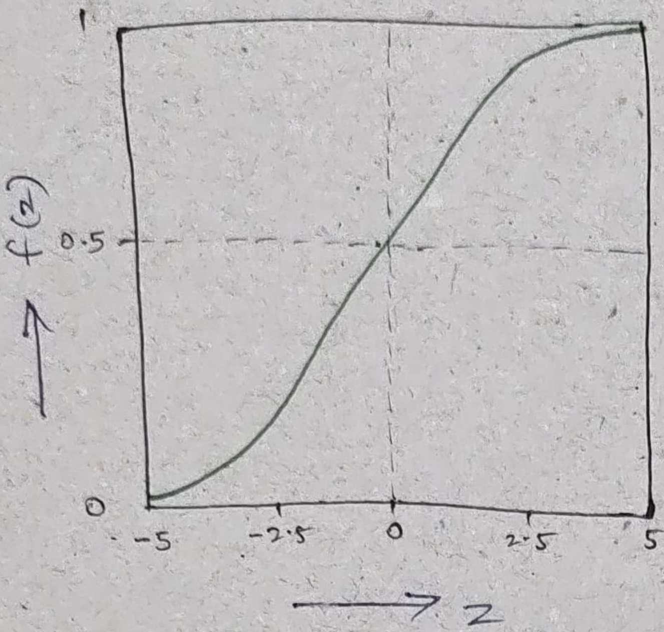
Also termed as the SIGMOID function.

Observe :

$$\lim_{z \to -\infty} \frac{1}{1+e^{-z}} = 0$$

$$\lim_{z \to \infty} \frac{1}{1+e^{-z}} = 1$$

Plot of Logistic function is below.



Probabilities

$$P(y=1|\bar{x}) = \frac{1}{1+e^{\underbrace{-\bar{x}^T\bar{h}}_{\text{inner product}}}} = g(x)$$

feature

$$P(y=0|\bar{x}) = 1 - P(y=1|\bar{x})$$

$$= 1 - \frac{1}{1+e^{-\bar{x}^T\bar{h}}}$$

$$= \frac{1+e^{-\bar{x}^T\bar{h}} - 1}{1+e^{-\bar{x}^T\bar{h}}}$$

$$= \frac{1}{1+e^{\bar{x}^T\bar{h}}}.$$

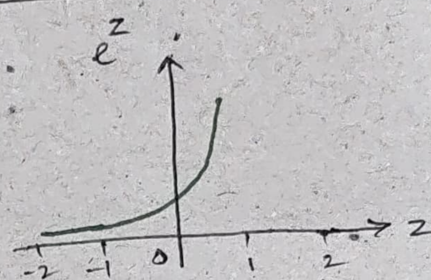
Note:

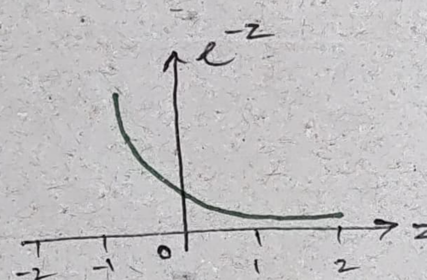
$$\lim_{z\to\infty} e^{z} = \infty$$

$$\lim_{z\to-\infty} e^{z} = 0$$

$$\lim_{z\to\infty} e^{-z} = 0$$

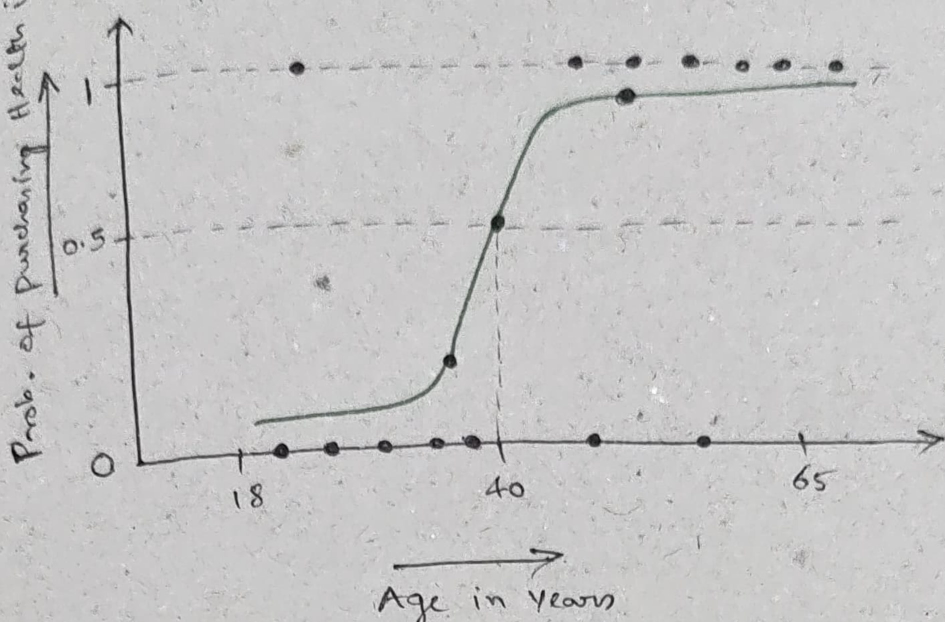
$$\lim_{z\to-\infty} e^{-z} = \infty$$





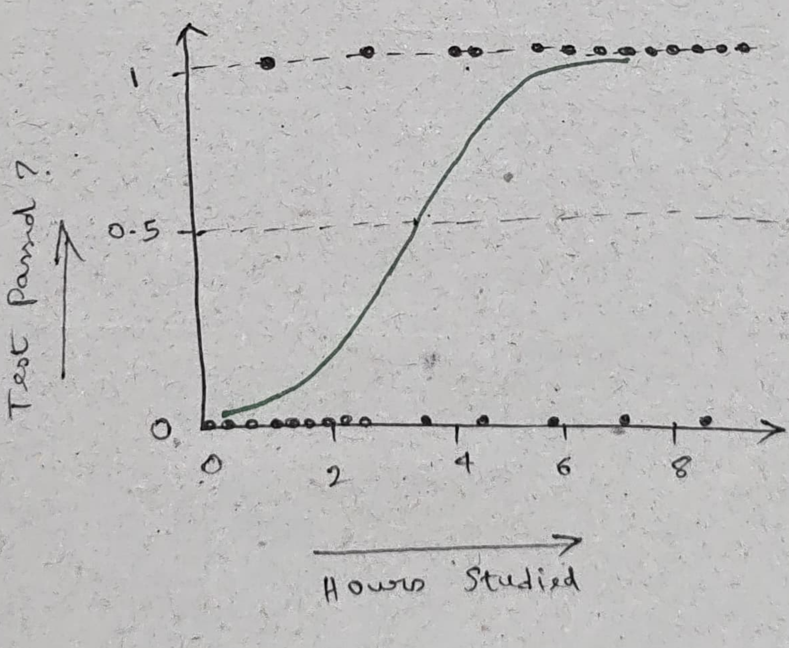
Examples :

(i) P (Health insurance | Age)



- Probability < 0.5 signifies that people less than 40 years are not likely to purchase insurance
- Threshold value 0.5 signifies that people having age more than 40 years are more likely to purchase insurance.
- Probability > 0.5 means, people older than 40 years are going to purchase health insurance.

(ii) P (Pass | Hours studied)



How to determine the regression parameter $\bar{h}$ in this case ?

We use the Maximum Likelihood technique !

- The Likelihood (i) Joint probability of $(y(k), \bar{x}(k))$ can be written as

$$\left(g\left(\bar{x}(k)\right)\right)^{y(k)} \left(1 - g\left(\bar{x}(k)\right)\right)^{1-y(k)}$$

- The Joint likelihood of all outputs / responses is

$$L(\bar{h}) = \prod_{k=1}^{M} \left(g\left(\bar{x}(k)\right)\right)^{y(k)} \left(1 - g\left(\bar{x}(k)\right)\right)^{1-y(k)}$$

- The Log likelihood is given as

$$\ln L(\bar{h}) = \ell(\bar{h})$$

$$= \sum_{k=1}^{M} y(k) \ln g\left(\bar{x}(k)\right) + (1-y(k)) \ln\left(1 - g\left(\bar{x}(k)\right)\right)$$

- To <u>maximize</u> the log-likelihood, one can employ <u>gradient ascent</u>.

<u>Logistic update rule</u>

Note that $f(z) = \frac{1}{1+e^{-z}}$

$$\frac{d}{dz} f(z) = \frac{e^{-z}}{(1+e^{-z})^2}$$

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

$$= f(z)\left(1 - f(z)\right).$$

The gradient of the likelihood can be evaluated as follows.

$$\nabla_{\ell}(\bar{h}) = \nabla\left(y(k)\ln g(\bar{x}(k)) + (1-y(k))\ln\left(1-g(\bar{x}(k))\right)\right)$$

$$= \frac{d}{d\,g(\bar{x}(k))}\left(y(k)\ln g(\bar{x}(k)) + (1-y(k))\ln\left(1-g(\bar{x}(k))\right)\right)$$

$$\times\ \nabla_{g}(\bar{x}(k))$$

$$= \left(\frac{y(k)}{g(\bar{x}(k))} - \frac{(1-y(k))}{1-g(\bar{x}(k))}\right) \times \nabla_{f}(\bar{h}^{T}\bar{x}(k))$$

$$= \left(\frac{y(k)}{g(\bar{x}(k))} - \frac{(1-y(k))}{1-g(\bar{x}(k))}\right)$$

$$\times f(\bar{h}^{T}\bar{x}(k)) \times \left(1-f(\bar{h}^{T}\bar{x}(k))\right) \times \bar{x}(k)$$

$$= \left(\frac{y(k)}{g(\bar{x}(k))} - \frac{(1-y(k))}{1-g(\bar{x}(k))}\right) g(\bar{x}(k))$$

$$\times\left(1-g(\bar{x}(k))\right)\bar{x}(k)$$

$$= \left(y(k) - g(\bar{x}(k))\right)\bar{x}(k)$$

The Update equation is given as

$$\bar{h}(k) = \bar{h}(k-1) + \eta\,\nabla_{\ell}(\bar{h})\Big|_{\bar{h}=\bar{h}(k-1)}$$

$$= \bar{h}(k-1) + \eta\left(y(k) - g(\bar{x}(k))\Big|_{\bar{h}=\bar{h}(k-1)}\right)\bar{x}(k)$$

$$= \bar{h}(k-1) + \eta\,\underline{\underline{e(k)}}\,\bar{x}(k)$$

$$e(k) = y(k) - g(\bar{x}(k))\Big|_{\bar{h}=\bar{h}(k-1)}$$

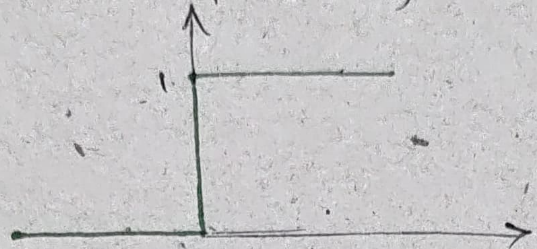
Least Mean Square

observe this is similar to the LMS update rule!

In <u>Perceptron Learning algorithm</u>, g is given as the threshold function

$$g(x) = \begin{cases} 1, & \bar{h}^T \bar{x} \geq 0 \\ 0, & \bar{h}^T \bar{x} < 0 \end{cases}$$

unit step (threshold)



The update rule (iterative way of computing solution) is once again given as

$$\bar{h}(k) = \bar{h}(k-1) + \eta \, e(k) \, \bar{x}(k)$$

$$e(k) = \underbrace{y(k)}_{\text{Actual o/p}} - \underbrace{g(\bar{x}(k)) \Big|_{\bar{h} = \bar{h}(k-1)}}_{\text{Predicted o/p}}$$

$$g(\bar{x}(k)) \Big|_{\bar{h} = \bar{h}(k-1)} = \frac{1}{1 + e^{-\bar{x}^T(k) \bar{h}(k-1)}}$$

This is termed as Perceptron Learning Algorithm.

It was proposed as a model for the <u>neurons</u> in the human brain.