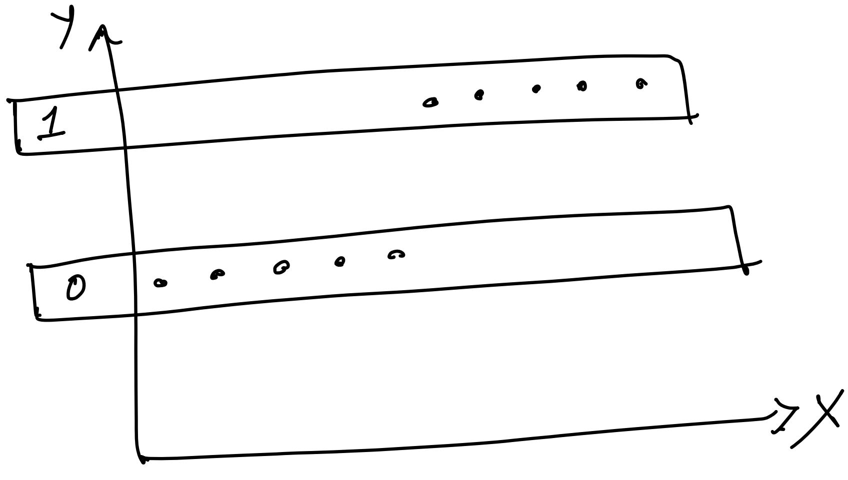
# Logistic Regression (LR)

→ Logistic regression is a special case of regression analysis and is used when dependent variable is nominally scaled.

Eg: Variable purchase decision : buys a product or does not buys a product.

→ With LR, it is now possible to explain the dependent variable or estimate the probability of occurrence of the categories of the variable.

→ In the basic form of LR, dichotomous variable (0 or 1) can be predicted. For this purpose, the probability of occurrence of value = 1 (characteristic present is estimated).




Age,
Gender,
Smoking status, → disease

$\underbrace{\text{Age, Gender, Smoking status}}$
↓
Independent variable

disease
↓
dependent variable
01 → diseased
00 → not diseased

(How likely is it that the disease is present if the person under consideration has a certain age, sex and smoking status).

<u>Calculate LR</u>:

To build a LR model, the linear regression eqn. is used as the starting point.

reg. co efficients

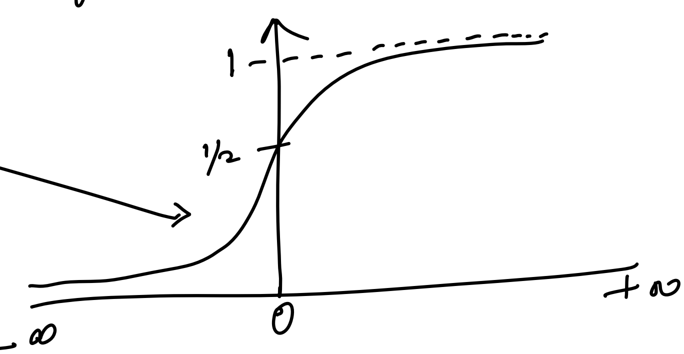
$$\hat{y} = b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + a$$

independent variable

<u>Logistic function</u> –

The logistic model is based on logical function.

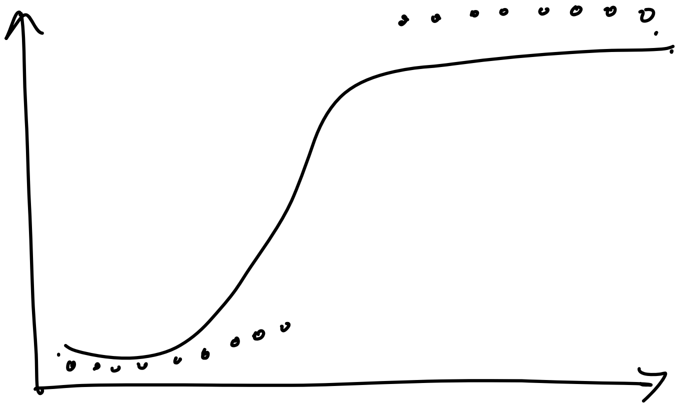
$$f(z) = \frac{1}{1+e^{-z}}$$




$$\hat{y} = b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + a.$$

$$f(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + a)}}$$

This now ensures that no matter in which range the $x$ values are located, only values between 0 and 1 will come out. The new graph now looks like this:



The probability that for given values of the independent variable the dichotomous dependent variable $y$ is 0 or 1 is given by:

$$P(y=1 \mid x_1, x_2, \dots, x_k) = \frac{1}{1+e^{-(b_1x_1+b_2x_2+\dots+b_kx_k+a)}}$$

$$P(y=0 \mid x_1\ x_2 \dots x_k) = 1-\frac{1}{1+e^{-(b_1x_1+\dots+b_kx_k+a)}}$$