UNIT - 2

Linear Regression

Regression → technique.

X = independent (Predictor) Y → dependent (target)

$y = f(x)$.

$y = f(x)$ if f is linear it is called linear Regression.

→ best fit straight line → regression line.

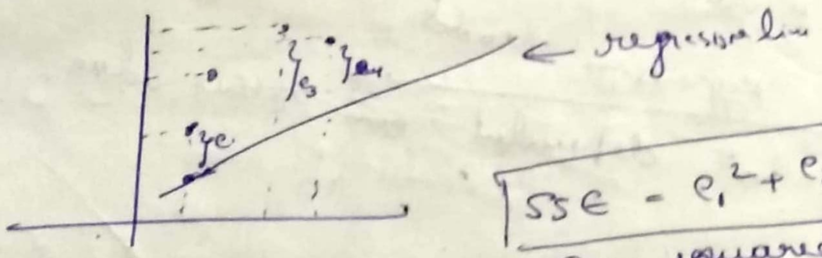
$y = \beta_0 + \beta_1 x$

$y_1 = \beta_0 + \beta_1 x_1$ ← we need to find β_0 & β_1

$y_2 = \beta_0 + \beta_1 x_2$. → Least square estimation method is used to find β_0 & β_1.

Simple Linear Regression (SLR).

→ Univariate → only 1 feature we used

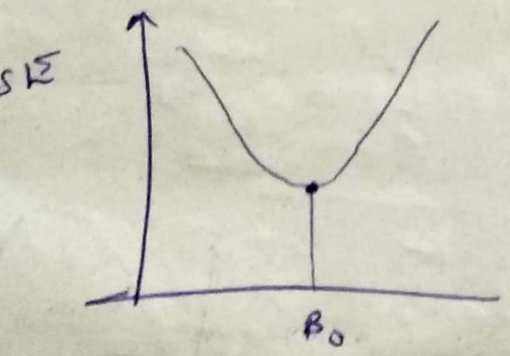


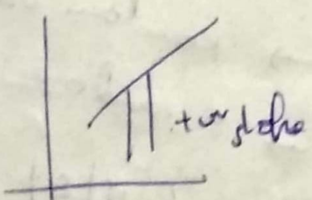
$$\boxed{SSE = e_1^2 + e_2^2 + e_3^2 + e_4^2}$$

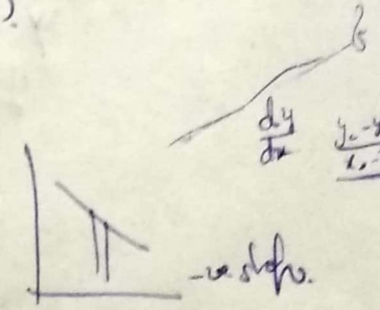
Sum squared error.

Least squared ~~er~~ estimate → select line which ~~the~~ / ~~has less SSE~~ minimum SSE.

(an optimization line)



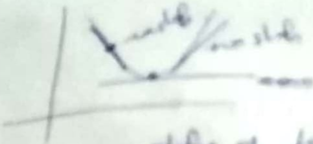




$\frac{dy}{dx}$ $\frac{y_2 - y_1}{x_2 - x_1}$

when we diff at any point we get ~~st~~ slope of tangent.

0 → mean not increasing or decreasing



slope of tangent at minimum point is zero

$\frac{\delta SSE}{\partial \beta_0}$

$SSE = \sum_{i}^{N} (y_i - \hat{y}_i)^2$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ → ^ indicates it's not exact.

$SSE = \sum_{i}^{N} \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)$

But $y = \hat{y} + e$

Random variable → a variable that have a probability of occurrence.

Univariate linear regression Model

cov = 0 i error is not dependent on other value

SLR is given as

$y = \hat{y} + e$

$y = \hat{\beta}_0 + \hat{\beta}_1 x + e$

y → is a random not Random variable

X → Random variable

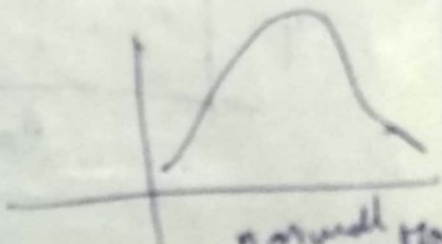
e → a random error

$E|\varepsilon| = 0$ & $V|\varepsilon| = \sigma^2$

mean $E(x)$ $V(x)$ variance

$cov|\varepsilon_i, \varepsilon_j| = 0$.

ε_i is normally distributed



normal distribution

Scanned by CamScanner

not

nowle

→ Once we fitted the data from gaussian distribution or standard distribution & we can find probability.

SSE function is convex function → proved fact

$$SSE = \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x)^2$$ to be minimize

we need to compute

$$\frac{\partial SSE}{\partial \beta_0} = 0 \quad \& \quad \frac{\partial SSE}{\partial \beta_1} = 0$$

$$\frac{\partial \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x)^2}{\partial \beta_0} = 0$$

$$\frac{\partial \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x)^2}{\partial \beta_1} = 0.$$

chain rule

$$\frac{dy}{dx} = \frac{du}{dz} \times \frac{dz}{du}$$

$$z = y_i - \beta_0 - \beta_1 x_i$$

$$\frac{\partial \sum_{i=1}^{N} z^2}{\partial \beta_0} = \frac{\partial \sum_{i=1}^{N} z^2}{\partial z} * \frac{\partial z}{\partial \beta_0}$$

$$\sum_{i=1}^{N} 2z * \frac{\partial z}{\partial \beta_0} \rightarrow 0 - 1 - 0)$$

$$\sum_{i=1}^{N} 2z * (-1)$$

$$\sum_{i=1}^{N} -2 (y_i - \beta_0 - \beta_1 x_i)$$

Scanned by CamScanner

$$\frac{\partial \sum_{i=1}^{N} z^2}{\partial \beta_1} = \frac{\partial \sum_{i=1}^{N} z^2}{\partial z} * \frac{\partial z}{\partial \beta_1}$$

$$= \sum_{i=1}^{N} 2z * (-x_i)$$

$$= \sum_{i=1}^{N} -2x_i (y_i - \beta_0 - \beta_1 * x_i)$$

$$\frac{\partial SSE}{\partial \beta_0} = -2 \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 * x_i) = 0 \quad \text{--- } (1)$$

$$\frac{\partial SSE}{\partial \beta_1} = -2 \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 * x_i) x_i = 0 \quad \text{--- } (2)$$

(1)

$$-2 \left[\sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 * x_i) \right] = 0$$

$$\sum_{i=1}^{N} y_i - \sum_{i=1}^{N} \underset{\uparrow \text{constant}}{\beta_0} - \sum_{i=1}^{N} \beta_1 x_i = 0$$

$$\sum_{i=1}^{N} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{N} x_i = 0$$

$$\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i = n\beta_0$$

$$\frac{1}{n} \sum_{i=1}^{n} y_i - \beta_1 \frac{1}{n} \sum_{i=1}^{n} x_i = \beta_0$$

$$\boxed{\bar{y} - \beta_1 \bar{x} = \beta_0}$$

$\bar{y}\ \bar{x} \to$ mean

Substitute β_0 in eqn (2)

$$\beta_1 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Q) Assume

X	0	1	2	3	4
Y	2	3	5	4	6

i) Find the least square regression line $Y = \beta_0 + X\beta_1$

ii) Plot the data + the line.

c) Estimate y for $x = 10$

d) compute SSE

$\Rightarrow$ $n = 5$

$$\beta_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x - \bar{x})^2}$$

$$\bar{X} = \frac{0+1+2+3+4}{5} = \frac{10}{5} = 2.$$

$$\bar{Y} = \frac{2+3+5+4+6}{5} = \frac{20}{5} = 4.$$

$$\beta_1 = \frac{(-2)(-2) + (-1)(-1) + (1)(0) + 0(1) + 2(2)}{(-2)^2 + 1^2 + 0^2 + 1 + 2^2}$$

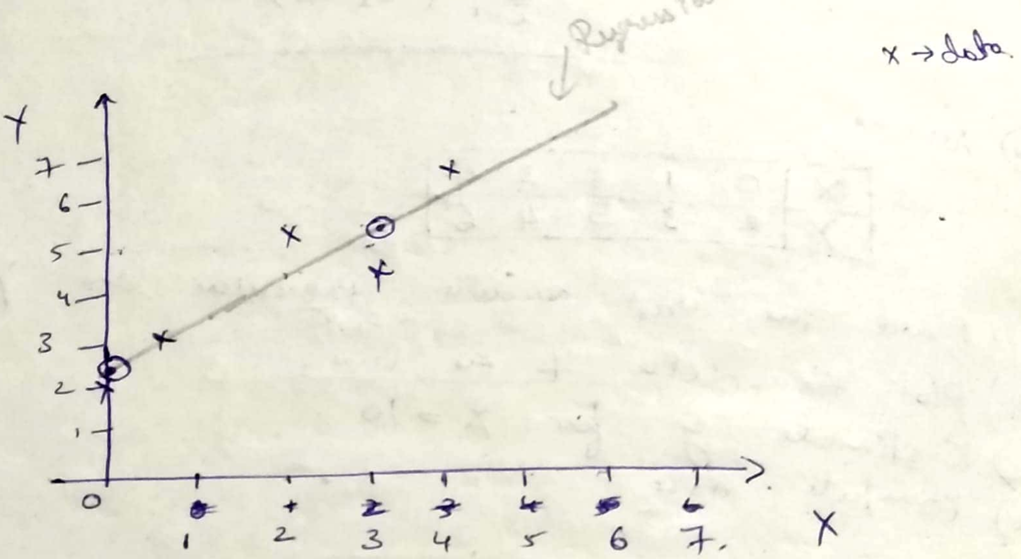
$$= \frac{4+1+0+0+4}{4+1+1+4} = \frac{9}{10} = \underline{\underline{0.9}}$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

$$= 4 - (0.9)2 = 4 - 1.8 = \underline{\underline{2.2}}$$

$\hat{Y} = 2.2 + 0.9x$ ← Regression line.

ii) Plot



$X_3 = (3, 4.9)$.

$y = 2.2 + 0.9(3)$
$= 2.2 + 2.7$
$= 4.9$

iii) $x = 10$ $\quad y = 2.2 + 0.9(10)$
$= 2.2 + 9 = \underline{\underline{11.2}}$

d). Compute SSE

~~SSE~~ ⇒ $\hat{y}_0 = 2.2 + 0.9(0) = \underline{\underline{2.2}}$

$\hat{y}_1 = 2.2 + 0.9(1) = \underline{\underline{3.1}}$

$\hat{y}_2 = 2.2 + 0.9(2) = \underline{\underline{4.0}}$

1.8

$\hat{y}_3 = 2.2 + 0.9(3) = 4.9$

$\hat{y}_4 = 2.2 + 0.9(4) = \underline{\underline{5.8}}$

3.6

$$SSE = (2-2.2)^2 + (3-3.1)^2 + (5-4.0)^2 + (4-4.9)^2 + (6-5.8)^2$$

$$= 0.2^2 + 0.1^2 + 1^2 + 0.9^2 + 0.2^2$$

$$= 0.04 + 0.01 + 1 + 0.81 + 0.04$$

$$1.90$$

$$\underline{SSE = 1.9}$$

(c)

x	-2	1	3
y	1	1	2

$$\bar{X} = \frac{-2+1+3}{3} = 2/3$$

$$\bar{Y} = \frac{1+1+2}{3} = 4/3$$

$$X_i - \bar{X} = -8/3,\ 1/3,\ 7/3$$

$$Y_i - \bar{Y} = -1/3,\ -1/3,\ 2/3$$

$$\beta_1 = \frac{(-8/3)(-1/3) + (1/3)(-1/3) + (7/3)(2/3)}{-8/3 + 1/3 + 7/3}$$

$$= \frac{8 - 1 + 14}{9} = \frac{21}{9}$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

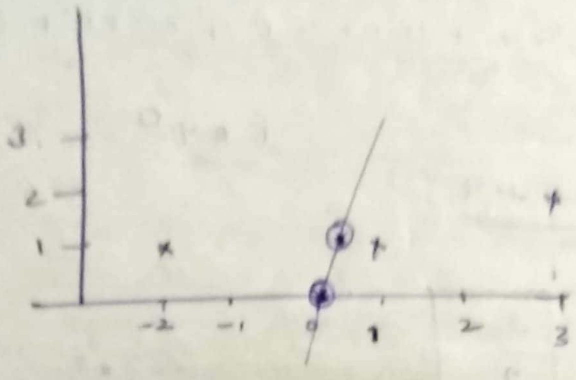
$$= 4/3 - \frac{21}{9}(2/3)$$

$$= 4/3 - \frac{14}{9} = \frac{-2}{9}$$

Scanned by CamScanner

Regression line = $-\frac{3}{9} + \frac{21}{9}x$

plot



Multivariate Regression (More than 1 o/p).
Multiple Regression. (more than 1 features)
↑X is multi dimensional.
$x_1 \quad x_2 \quad x_3 \quad x_4$.

$(x, y) \Rightarrow x \Rightarrow x_1, x_2 \dots x_p$.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p$$

error

$$y_i = \underbrace{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}_{\hat{Y}_i} + \varepsilon_i$$

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 X_{12} + \beta_3 X_{13} + \dots + \beta_p X_{1p} + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \beta_3 X_{23} + \dots + \beta_p X_{2p} + \varepsilon_2$$

$\vdots$

$$y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \beta_3 X_{n3} + \dots + \beta_p X_{np} + \varepsilon_2.$$

We add X_{i0} to β_0 to make it uniform

$$Y_1 = \beta_0 X_{10} + \beta_1 X_{11} + \dots + \beta_p X_{1p} + \varepsilon_1$$

$\beta_i X$

$\beta_0 X_0$

$$Y_n = \beta_0 X_{n0} + \beta_1 X_{n1} + \dots + \beta_p X_{np} + \varepsilon_n$$

$\sum_{0}^{p} \beta X$

every y

$$Y_i = \sum_{j=0}^{p} \beta_j X_{ij} + \varepsilon_i$$

Y is 1D vector

always

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y \\ y_n \end{bmatrix} = \begin{bmatrix} x_{10} & x_{11} & \cdots & x_{1p} \\ x_{20} & x_{21} & \cdots & x_{2p} \\ \vdots & & & \\ x_{n0} & x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$n \times 1$ $\qquad$ $\underbrace{n \times (p+1) \quad (p+1) \times 1}_{n \times 1}$ $\quad + \quad n \times 1$

$n \times 1 \quad + \quad n \times 1$

$$\boxed{Y = X\beta + \varepsilon}$$

$$SSE = \sum_{i=1}^{n} \varepsilon_i^2$$

$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 \end{bmatrix}$ $\quad \varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \dots$

$1 \times n$

$$\Rightarrow \varepsilon^T \varepsilon$$

$n \times 1$ $\quad n \times 1$ $\quad [\varepsilon_1 \; \varepsilon_2 \; \varepsilon_3] \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

$$= \frac{\partial SSE}{\partial \beta_p} = \frac{\partial SSE}{\partial \beta} = 0$$

driv w.r.to $\beta_0 + \beta_1 \dots \beta_p$

w.k $\quad \varepsilon = Y - X\beta$

$$\frac{\partial(\varepsilon^T\varepsilon)}{\partial\beta} = \frac{\partial (Y - X\beta)^T (Y - X\beta)}{\partial\beta}$$

$$= \frac{\partial \left(Y^T Y - Y^T X\beta - X^T\beta^T Y + X^T\beta^T X\beta \right)}{\partial\beta} = 0$$

dodndn

$$\beta = (X^T X)^{-1} X^T Y$$

$$Y = \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix} \qquad X = \begin{bmatrix} 5 \\ 7 \\ 10 \\ 12 \\ 20 \end{bmatrix} \leftarrow 10$$

Do multi variate LR.

$$X = \begin{bmatrix} 1 & 5 \\ 1 & 7 \\ 1 & 10 \\ 1 & 12 \\ 1 & 20 \end{bmatrix} \qquad Y = \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix} \qquad \beta = (X^T X)^{-1} X^T Y.$$

674

5×2

25 + 49 + 100 + 144 + 400

$$= \overset{X^T}{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 7 & 10 & 12 & 20 \end{bmatrix}} \overset{X}{\begin{bmatrix} 1 & 5 \\ 1 & 7 \\ 1 & 10 \\ 1 & 12 \\ 1 & 20 \end{bmatrix}}$$

$$= \begin{bmatrix} 5 & 54 \\ 54 & 718 \end{bmatrix}^{-1}$$

$$= \frac{1}{\cancel{2918}\ 674} \begin{bmatrix} 718 & -54 \\ -54 & 5 \end{bmatrix} = \begin{bmatrix} 1.07 & -0.08 \\ -0.08 & 0.007 \end{bmatrix}$$

$(X^T X)^{-1}$

if matrix is 2x2

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

determinant

718×5

$3590 -$

$$X^T Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 7 & 10 & 12 & 20 \end{bmatrix} \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix}$$

$$= \begin{bmatrix} 150 \\ 1970 \end{bmatrix}$$

$50 + 140 + 300 + 480 + 100.$

3025 −
2925

107

50
140
300
480
1000
970

$$\beta = (X^T X)^{-1} X^T Y$$

$$= \begin{bmatrix} 1.07 & -0.08 \\ -0.08 & 0.007 \end{bmatrix} \begin{bmatrix} 150 \\ 1970 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 2.9 \\ 1.79 \end{bmatrix} \Rightarrow \beta_0 = 2.9$$

$$\beta_1 = 1.79.$$

error

$$SSE = e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2$$

$y = \beta_0 + \beta_1 X$

$$e_1^2 = (y_1 - \hat{y}_1)^2.$$

$$= 10 - (2.9 + 1.79 * 5)$$

$$= (10 - 11.85)$$

$$e_1^2 = 1.85^2 = \underline{\underline{3.4225}}$$

Scanned by CamScanner

$\beta_0 = 2.9$,

$\beta_1 = 1.79$.

$e_2^2 = y_2 - \hat{y}_2$

$= 20 - 15.43 = 4.57$

$= \underline{20.8849}$

$e_3^2 = (30 - 20.8)^2$

$= 84.64$

$e_4^2 = 40 -$

$= \underline{243.98}$

$e_4^2 = 127.69$

$SSE = \underline{480.617}$

Regularized Regression

→ used to avoid overfitting.

→ It shrinks coefficient estimate towards zero.

→ discourages learning a complex method.

we try to minimize expression

β must be

extra

we are not overfitted

Ridge Regression

Loss function $= \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j * x_{ij})^2 + \lambda \sum_{j=1}^{p}$

like SSE (cost function).

we minimize it.

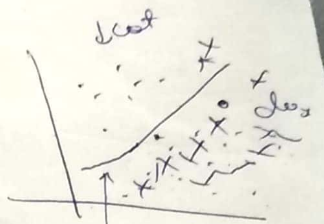
Lasso Regression

Loss function $= \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j * x_{ij})^2 + \lambda \sum_{j=1}^{p} |$

Scanned by CamScanner

* To min overall function,
if λ is more β must be less $\Rightarrow$ overall loss function is less
→ overfitting reduces
→ but it might be underfitted

Logistic Regression → actually it's a classification



cat $y = 1$
dog $y = 0$.

$P(Y=1/X) = 0.5$

$P(Y=1/X) > 0.5$ → cat

$P(Y=1/X) < 0.05$ → dog.

linear regression with $p(x) = \beta_0 + \beta_1 x$ ← cat be used
we get Y.
But here in logistic Regression we want out as probability.
→ probability has to be 0 & 1. Therefore we apply log.

$$\log\left[\frac{p(x)}{1-p(x)}\right]$$

note $p(x) = p(y=1/x)$

$$\log\left[\frac{p(x)}{1-p(x)}\right] = \beta_0 + \beta_1 x$$ ← value will be [0

$$\frac{p(x)}{1-p(x)} = e^{(\beta_0 + x \cdot \beta_1)}$$

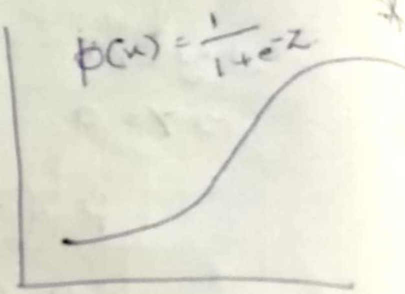
$$p(x) = (1 - p(x))\, e^{\beta_0 + x \cdot \beta_1}$$

$$p(x) = \frac{e^{\beta_0 + x \cdot \beta_1}}{1 + e^{-(\beta_0 + x \cdot \beta_1)}}$$ ← sigmoid function

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + x\beta_1)}}$$

use this to compute

logistic function sigmoid



Multiple Logistic Regression

for many $\beta_0 + \beta_1$ same formula

if we derive

$$\frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right)$$

$$g(z)\,(1 - g(z))$$

easy derivative
simple derivate
1 of its advantage

Maximum likelihood estimation.

Cost function ←

~~dec~~

$$l(\beta_0, \beta_1) = \prod_{i: y_i = 1} p(x_i) \prod_{i': y_{i'} = 0} 1 - p(x_{i'})$$

where $y = 1$

$p(Y=1|x)$ 1

0

1

so its max.

Artificial Neural Network

→ derived from brain
→ N.N exist we try to find a algorithm, which behaves like that.

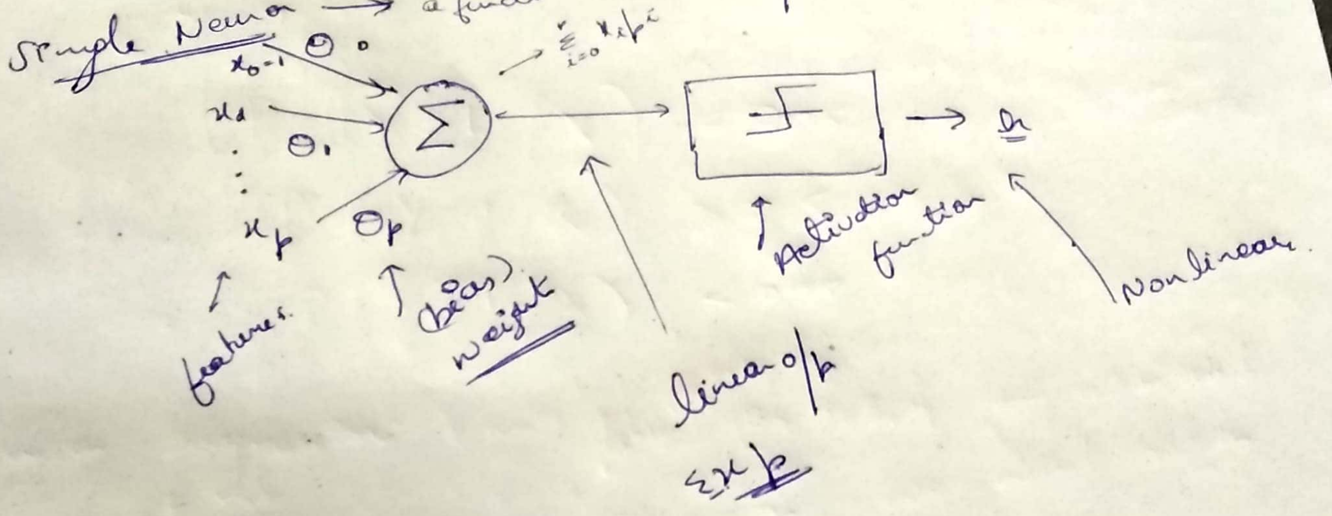
UNIT-2

Linear classification → line separate classes.

Non linear classification → we need to draw a curve according

→ ANN can solve it.

Simple Neuron → a function occur I/p's



2 functionality

1) Summation
2). Activation → linear $f(x) = 1$

threshold → Nonlinear

sigmoid → Nonlinear