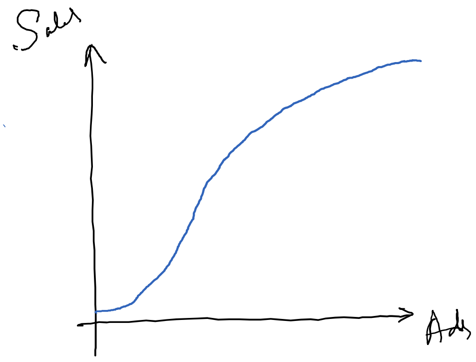
Machine Learning



Correlation → Relationship btn 2 variables.

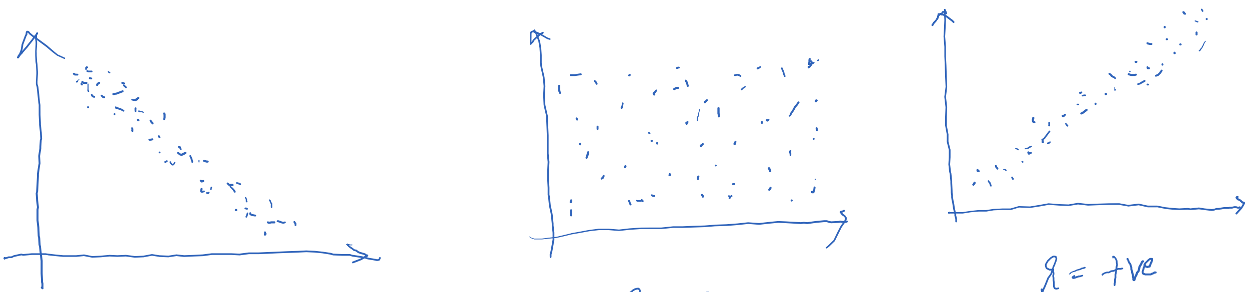
ρ → Population → Not possible in Reality

r → Sample-Corelation Co-efficient

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \cdot \sum (y-\bar{y})^2}}$$

→ Pearsons Co-relation Co-efficient

$[-1 \text{ to } +1]$



$r = -ve$ $r = 0$ $r = +ve$

Regression → Constructing a mathematical model that can predict one variable by using other Variables

one variable → Dependent

other Variables → Independent

Simple linear regression

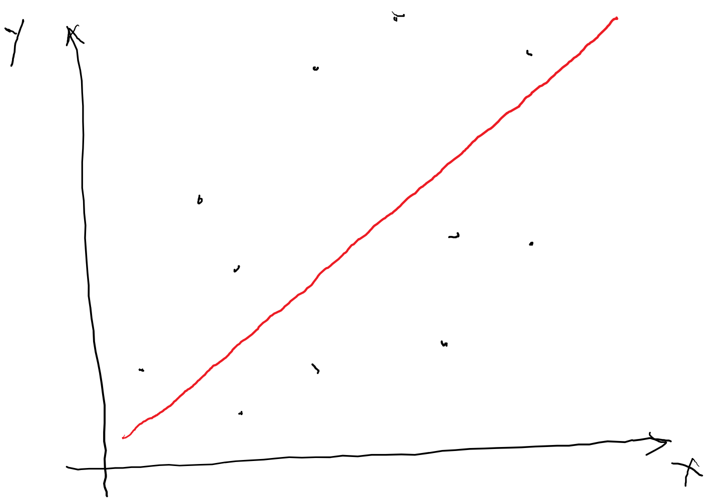
$$y = mx + c$$

y → Dependent

m → Slope

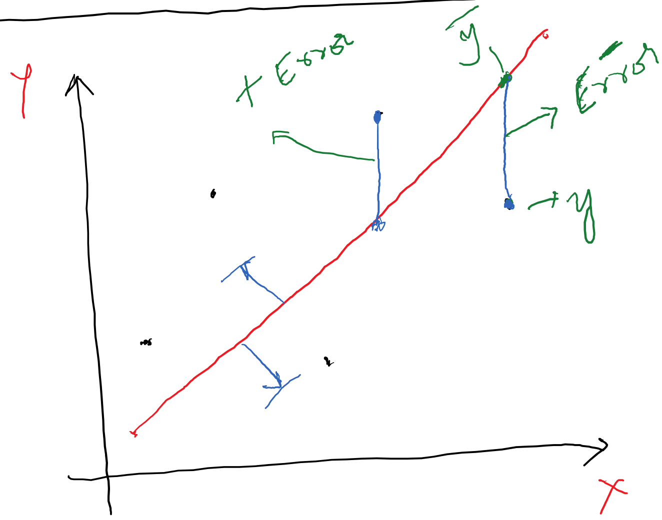
x → Independent var

c → Intercept-y



Step 1 → plot → scatterplot

for a sample ⇒ $\hat{y} = b_0 + b_1 x$



→ Determine the equation of regression line

Least square analysis

Residual/Error = $(y - \hat{y})$

$\Sigma (y - \hat{y}) \simeq 0$

⇒ Sum of squared residuals = $\Sigma (y - \hat{y})^2$ → limited use → Square root
SSE → Adjust for n.

→ Standard error of estimation = $S_e = \sqrt{\frac{SSE}{n-2}}$

Outlier → Data points that deviate very much from the mainstream data

68% → error terms → $1 S_e$
95% —— $2 S_e$
99.7% —— $3 S_e$
Outlier

Co-efficient of Determination → (Explained Variance of Model) → 80%

R^2

$SS_{yy} = SSR + SSE \quad \div SS_{yy}$

$1 = \boxed{\frac{SSR}{SS_{yy}}} + \frac{SSE}{SS_{yy}}$

$\boxed{R^2 = 1 - \frac{SSE}{SS_{yy}}}$

Linear reg
R → Co-relation
R^2 → Determination
$R^2 = (R)^2$

$$\to \boxed{r^2} = 1 - \frac{SSE}{SS_{yy}}$$

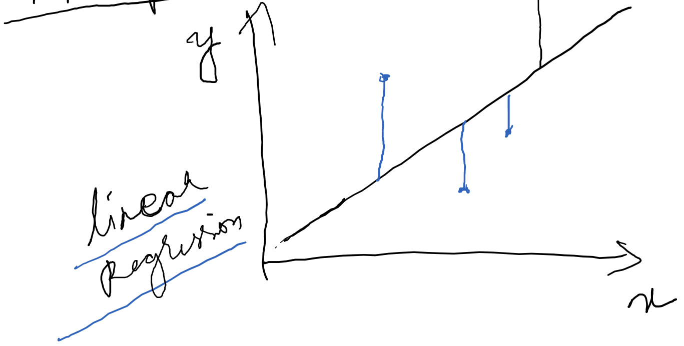
$r \to$

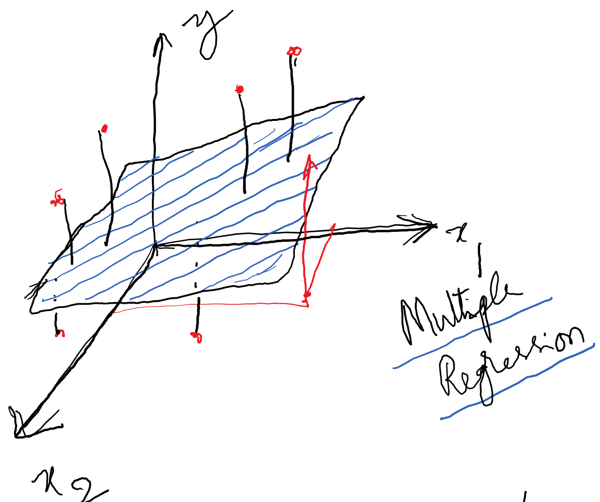
$r^2 = (r)^2$

$r \to$ How much y changes for one variation in x

$r^2 \to$ How much Variance in y is explained by x

Multiple Regression





$$y = mx + C$$

$$\hat{y} = b_0 + b_1 x \quad \longrightarrow \quad \hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k$$

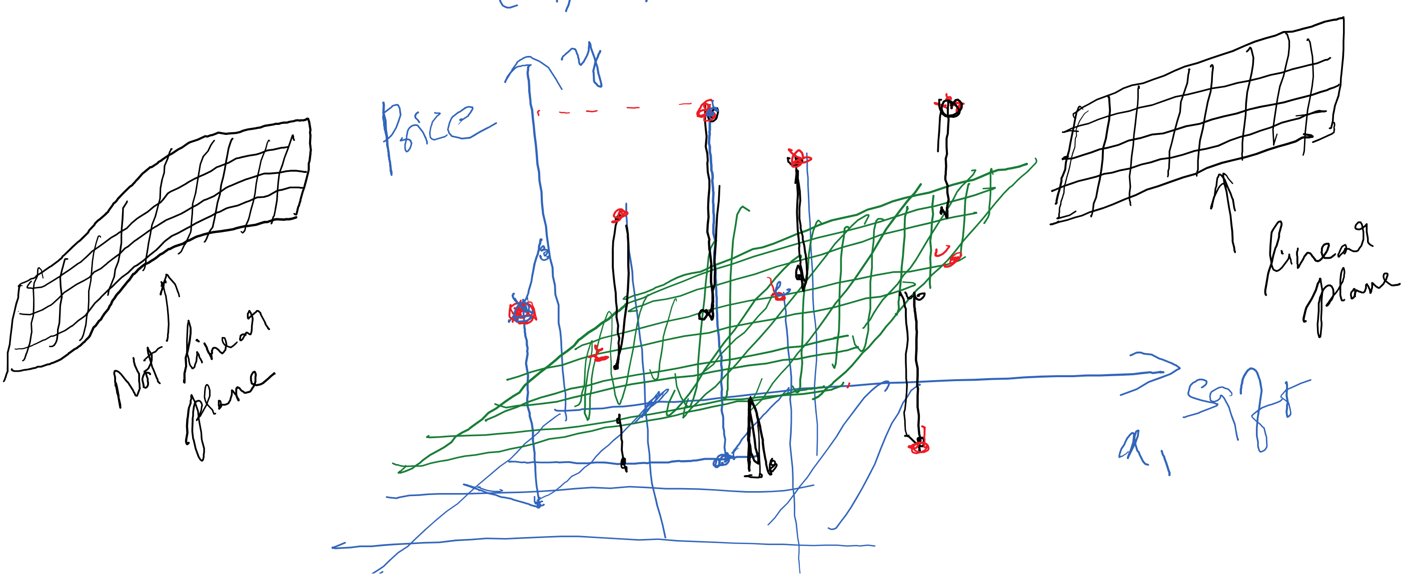
$\to$ x & y axis

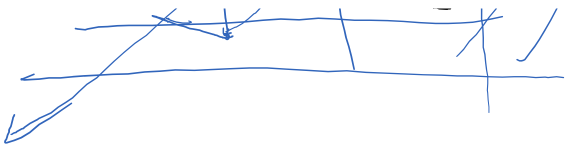
$\to$ For simplicity we use 2 - Independent & 1 Dependent var

$$\Rightarrow \hat{y} = b_0 + b_1 x_1 + b_2 x_2$$

(x, y, z)

$\Rightarrow (x_1, x_2, \hat{y})$





Age x y

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$

- $\hat{y}$ → Dependent Var
- b_0 → y Intercept
- x_1 → Indp (1)
- x_2 → Indp (2)

b_1 → Estimated Value of regression Co-efficient 1

b_2 → Estimated Value of regression Co-efficient 2

b_0 → Estimated Value of Regression Constant.

b_i → Represents the change that will occour in the Value of "y" for unit change in (x_i) that independent variable

⊛ If all other Variables are Constant ⊛

$$y = b_0 + b_1 \underbrace{x_1} + b_2 \underbrace{x_2} + b_3 x_3$$

$$y = 1 + \underbrace{3 \quad 2,3}_{\substack{b_i \quad x_i \\ i=1 \quad i=1}} + \underbrace{100 \quad 2,3}_{\substack{b_i \quad x_i \\ i=2 \quad i=2}} + 5 \; x_3$$

3 [207
  [210

Residual / Error → $(y - \bar{y})$ $\qquad \Sigma(y - \bar{y}) \simeq 0$

$SSE = \sum (y - \bar{y})^2$ → Minimum for a good fit plane

↳ square units

R^2 → Coefficent of determination $\neq (r)^2$ ← (Square root)

R^2 → Proportion of Variance of dependent variable occouring for the respective independent Variable for the respective independent variable

→ Multiple R^2

$$s_e = \sqrt{\frac{SSE}{n-2}}$$

→ Accuracy

→ % Explained Variance

Add

⟨+1 var⟩

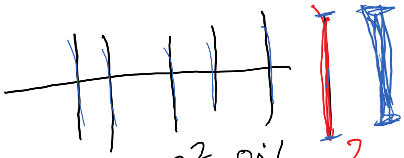
+1 var

⋮

+k^{th} var

Adjusted R^2

$$Adj\,R^2 = 1 - \frac{SSE/(n-k-1)}{SS_{yy}(n-1)}$$



$R^2 = 80\%$

$R^2 \downarrow 70\%$

$R^2 \uparrow 85\%$

$R^2 \Rightarrow$ Explained

Exp Var ↑ Unexp Var ↑

$$\boxed{SS_{yy} = SSR + SSE}$$

SS_{yy} ÷

$$\frac{SS_{yy}}{SS_{yy}} = \underbrace{\boxed{\frac{SSR}{SS_{yy}}}}_{R^2} + \frac{SSE}{SS_{yy}}$$

$$\Rightarrow 1 = R^2 + \frac{SSE}{SS_{yy}}$$

$$\boxed{R^2 = \frac{SSR}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}}}$$