In Earlier Example we had one Dependent variable ie y
and one Independent variable ie x
So it was case of Simple linear Regression.

Ex sqftarea → price ⇒ $\hat{y} = \beta_0 + \beta_1 x$

But if

x_1 Size (feet)^2
x_2 No of Bedrooms
x_3 No of floor
x_4 Location

Price
y

← multiple linear Regression.

The linear Regression to express the above dependency is

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

General form of Multiple linear Regression.

$$\boxed{\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n}$$

Here $\beta_0, \beta_1, \beta_2 \ldots \beta_n$ < +ve OR -ve

$$\boxed{h_\theta(x) = \beta_0 \underline{x_0} + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}$$

$x_0 = 1$

Here $x_0, x_1, x_2 \dots x_n$ = features

$\beta_0, \beta_1, \beta_2 \dots \beta_n \Rightarrow$ Parameter / coefficients.

Let X = Feature Vector = $\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}_{n+1 \times 1}$

$\underline{\underline{n+1 \times 1}}$

n = no of features

θ = Parameter Vector = $\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$ $n+1 \times 1$

$$\therefore \boxed{h_\theta(x) = \theta^T . X} = \underbrace{\begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \dots & \beta_n \end{bmatrix}}_{\theta^T} * \underbrace{\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{X}$$

$$\underline{\underline{h_\theta(x)}} = \theta^T X = \underline{\beta_0 x_0 + \beta_1 x_1 + \dots + \beta_n x_n}.$$

det x_0	X_1	X_2	y
1	7	2.6	78.5
1	1	2.9	74.3
1	11	5.6	104.3
1	11	3.1	87.6
1	7	5.2	95.9
1	11	5.5	109.2
1	3	7.1	102.7

3 2 2

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

To find β_0, β_1, β_2

Solve ->

Feature Vector = $X = \begin{bmatrix} 1 & 7 & 2.6 \\ 1 & 1 & 2.9 \\ 1 & 11 & 5.6 \\ 1 & 11 & 3.1 \\ 1 & 7 & 5.2 \\ 1 & 11 & 5.5 \\ 1 & 3 & 7.1 \end{bmatrix}_{7\times3}$ & $y = \begin{bmatrix} 78.5 \\ 74.3 \\ 104.3 \\ 87.6 \\ 95.9 \\ 109.2 \\ 102.7 \end{bmatrix}_{7\times1}$

$$\boxed{C = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = (X^T . X)^{-1} . X^T . y}$$

coefficient vector

3x1

Solution

$X = \begin{bmatrix} 1 & 7 & 2.6 \\ 1 & 1 & 2.9 \\ 1 & 11 & 5.6 \\ 1 & 11 & 3.1 \\ 1 & 7 & 5.2 \\ 1 & 11 & 5.5 \\ 1 & 3 & 7.1 \end{bmatrix}$

(1) find $X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 7 & 1 & 11 & 11 & 7 & 11 & 3 \\ 2.6 & 2.9 & 5.6 & 3.1 & 5.2 & 5.5 & 7.1 \end{bmatrix}$

(2) find $\underset{3\times7\ \ 7\times3}{X^T . X} \Rightarrow \begin{bmatrix} 7 & 51 & 32 \\ 51 & 471 & 235 \\ 32 & 235 & 163.84 \end{bmatrix}_{3\times3}$

(3) find $(X^T . X)^{-1}$

$= \begin{bmatrix} 1.79 & -0.06 & -0.25 \\ -0.06 & 0.01 & -0.0011 \end{bmatrix}$

$$= \begin{bmatrix} 1.79 & -0.06 & -0.25 \\ -0.06 & 0.01 & -0.0011 \\ -0.25 & -0.0011 & 0.0571 \end{bmatrix}$$

$$= (X^T \cdot X)^{-1} \cdot X^T \cdot y$$

$$\theta = \begin{bmatrix} 51.6 \\ 1.5 \\ 6.72 \end{bmatrix}_{3 \times 1} \qquad \begin{aligned} \beta_0 &= 51.6 \\ \beta_1 &= 1.5 \\ \beta_2 &= 6.72 \end{aligned}$$

$$\boxed{y = 51.6 + 1.5x_1 + 6.72x_2}$$

what is y when $x_1 = 3$ $x_2 = 2$

$$y = 51.6 + (1.5 \times 3) + 6.72 * 2 = 69.54$$

Consider Formulae that will be used, when only 2 Independent Variables specified.

[>2 features, Use Matrix Algebra].

Consider

x_1	x_2	y
3	8	-3.7
4	5	3.5
5	7	2.5
6	3	11.5
2	1	5.7
3	2	?

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$\theta_0 = ?$

$\theta_1 = ?$

$\theta_2 = ?$

$$\theta_0 = \bar{y} - \theta_1 \bar{x_1} - \theta_2 \bar{x_2}$$

$\left.\begin{array}{l} \bar{y} = \\ \bar{x}_1 = \\ \bar{y}_1 = \end{array}\right\}$ mean

$$\theta_1 = \frac{(\Sigma x_2^2)(\Sigma x_1 y) - (\Sigma x_1 x_2)(\Sigma x_2 y)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

$$\theta_2 = \frac{(\Sigma x_1^2)(\Sigma x_2 y) - (\Sigma x_1 x_2)(\Sigma x_1 y)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

where

$$\Sigma x_1^2 = \Sigma x_1 x_2 - \frac{(\Sigma x_1)(\Sigma x_1)}{N}$$

$$\Sigma x_2^2 = \Sigma x_2 x_2 - \frac{(\Sigma x_2)(\Sigma x_2)}{N}$$

$$\Sigma x_2 = \Sigma x_2 x_2 - \frac{(\Sigma x_2)(\Sigma x_2)}{N}$$

$$\Sigma x_1 x_2 = \Sigma x_1 x_2 - \frac{(\Sigma x_1)(\Sigma x_2)}{N}$$

$$\Sigma x_1 y = \Sigma x_1 y - \frac{(\Sigma x_1)(\Sigma y)}{N}$$

$$\Sigma x_2 y = \Sigma x_2 y - \frac{(\Sigma x_2)(\Sigma y)}{N}$$

Sol^n = $\boxed{y = 2.796 + 2.28 x_1 - 1.67 x_2}$

HW

* To Evaluate Performance of M.L models
* To find how good the model fits on given data

① Karl Pearsons Coefficient of Correlation (r)

→ To calculate Relationship bet two variable

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{\left(N*\sum x^2 - (\sum x)^2\right)\left(N*\sum y^2 - (\sum y)^2\right)}}$$

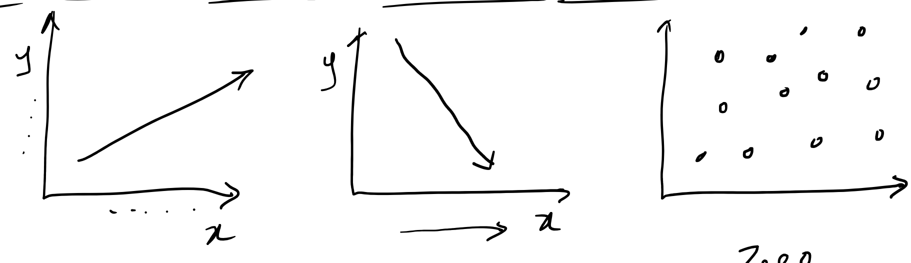
r = quantifies the strength of relationship bet^n two variables.

The value of r be between $+1$ and -1

If $r = 1$ ⇒ Total +ve correlation ⇒ If $x\uparrow$ then $y\uparrow$

If $r = -1$ ⇒ Total -ve correlation ⇒ If $x\downarrow$, then $y\uparrow$ or $x\uparrow$ then $y\downarrow$

→ It gives strength (degree) & direction of correlation.



+ve Correlation — -ve Correlation — Zero Correlation.

Q. H.W

x	y
151	63
174	81
138	56
186	91
128	47
136	57

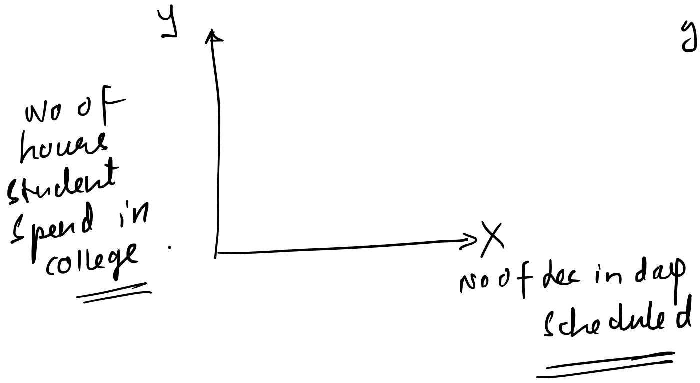
Find r = ? 0.9884

186	91
128	47
136	57
179	76
163	72
152	62
131	48

② R^2 method (R Square)

↳ It gives information about good of fit feature of the model.

↳



If $r^2 = 0.85$

Variation in No of hours that students spend in college is 85% dependent on No of lec scheduled.

↳ Indicate percentage of variance in dependent and Independent variable pair

↳ value varies from 0 to 1

If $r^2 = 1$ ⇒ no diff betⁿ actual & predicted value.

$r^2 = 0$ means the model does not learn any relationship betⁿ variables.

① SST (Sum of squares of Total) = $\Sigma (y - \bar{y})^2$

actual y, mean of y.

② SSR (Sum of Squares due to Regression) = $\Sigma (\hat{y} - \bar{y})^2$

predicted y, mean of y.

③ SSE (Sum of Squares of Error) = $\Sigma (y - \hat{y})^2$

actual y, predicted y

$$R^2 = SSR/SST = \boxed{\frac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2}}$$

Q.

H.W

x	y	$\hat{y}$
151	63	
174	81	
138	56	
186	91	
128	47	
136	57	
179	76	
163	72	
152	62	
131	48	

To find R^2

To find Predicted i.e $\hat{y}$

$$y = \beta_0 + \beta_1 x$$

find $\beta_0 = -38.455$

$\beta_1 = 0.6746$

$$\boxed{y = -38.455 + 0.6746x}$$

Now $x = 151$ find $\hat{y} = 63.4$

$x = 174$ $\hat{y} = 78.92$

$x = 138$ $\hat{y} = 54.63$:

Note $R^2 = \dfrac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2}$

if line fits properly

then $\hat{y} \simeq y$

$\therefore$ $R^2 \simeq 1$

(3) Standard Error of Estimate →

* Measures the accuracy of Prediction.

* $\underline{\sigma_{est}} = \sqrt{\dfrac{\Sigma(y-\hat{y})^2}{N}} = \sqrt{MSE} = \sqrt{\dfrac{SSE}{N}}$

↑ (mean Square Error)

y = actual value
$\hat{y}$ = predicted value.

* It reflects how well the Regression model fits the dataset.

* Smaller the value better it is
* Larger the value worst it is

Q.

H.W

x	y
151	63
174	81
138	56
186	91
128	47
136	57
179	76
163	72
152	62
131	48

$$\boxed{\sigma_{est} = 2.909}$$

Classification →

Ex Mail → Spam (+ve) / Not Spam (-ve)

Online Transaction → Fraudulent (+ve) / Non Frodulent (-ve)

Tumor → Malignent (+ve) / Benign (-ve)

where $y \in \{0, 1\}$
- 0 ⇒ -ve class
- 1 ⇒ +ve clas.

Here y has Discrete Value.

In above case o/p has only 2 class (0, 1) } Binary classification.

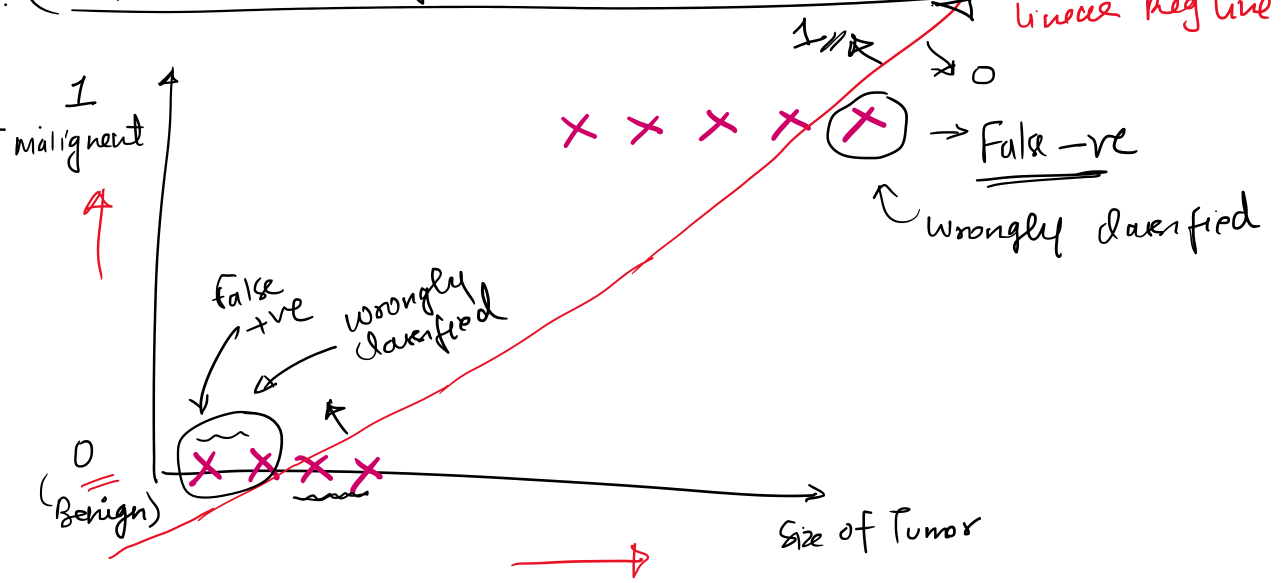
To check Weather.

Possible o/p: Windy, Sunny, Cloudy, Rainy } o/p has more than one class [multi class Classification]

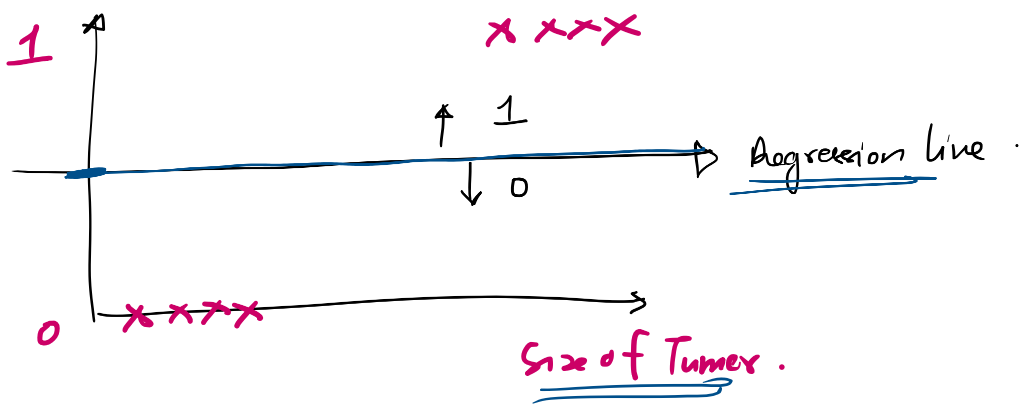
* Classification bothers about label and not the Exact value.

Consider (Why Linear Regression is not used for Classification)

Consider



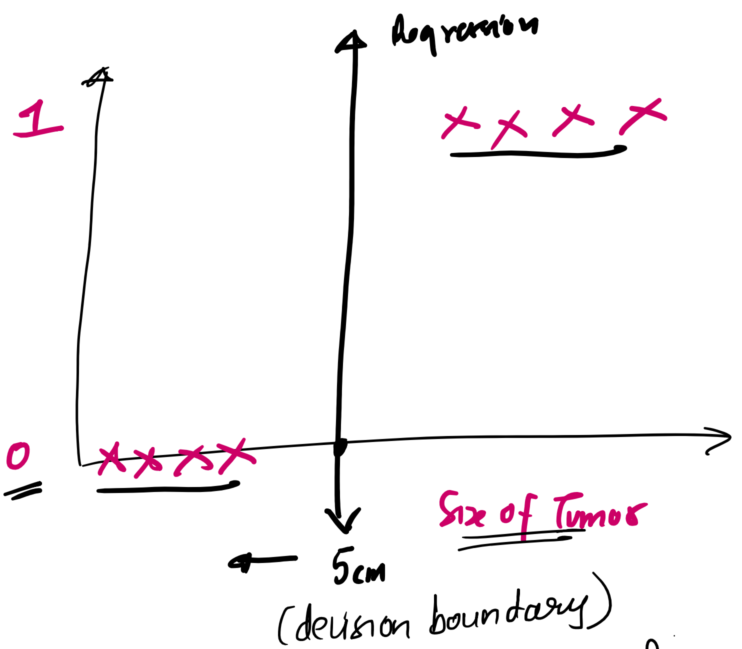
Suppose



malignancy

Not Correct

Consider



Malignancy.

Here
we can observe for above Regression line

if tumor size ≤ 5cm → Yes (1) Malignant

If tumor size $\leq$ 5cm $\rightarrow$ Yes (1) Malignant
tumor size > 5m $\longrightarrow$ No (0) Benign

We can say let 'P' denote Probability that $y=1$ when $X=x$.

$$y = P_r(y=1 \mid X=x) = [\beta_0 + \beta_1 x]$$

In linear Reg

P = probability lies betn 0 to 1

But linear function are unbounded.
and Expected o/p here is 0 or 1
So we cannot use Regression to build Classifier

$\therefore$ Linear Regression is not suitable for Classification.

For Classification we will use Logistic Regression.

* In Logistic Regression we get probability score.
* It predicts the probability of occurance of event

$$\text{Odd} = \frac{\text{No of time the Event happens}}{\text{No of time the Event will not happen}}$$

[Represents chances that the event will occur]

Ex If odd of India winning against W.I is 4:1 $= \frac{\text{No of India win}}{\text{No of India not win}}$

$= \frac{4}{1}$

Best Case ⇒ The odd of India winning against WI = ∞

If odd of W.I winning against India is 1:4

$= \frac{\text{No of W.I win}}{\text{No of W.I not win}} = \frac{1}{4}$

Worst Case ⇒ W.I is winning 0 match = Odd = 0

Range of value that odd can take = 0 to ∞

Relationship between Odd & Probability = $\text{Odd} = \frac{P}{1-P}$