2 Jan, 19

Machine Learning

feeding machine with necessary data to make it learn.

There are 2 types of machine learning techniques

Supervised Learning / predictive Learning	Unsupervised learning / Descriptive Learning
- We have a trainer, teacher to sort/ classify samples to a particular type.	- no supervisor self classify objects into some patterns.
- eg:- classification of fishes in a pond in a particular class label.	

NOTE:- Parameters should be such $y = \hat{y}$ {ideal condition}.

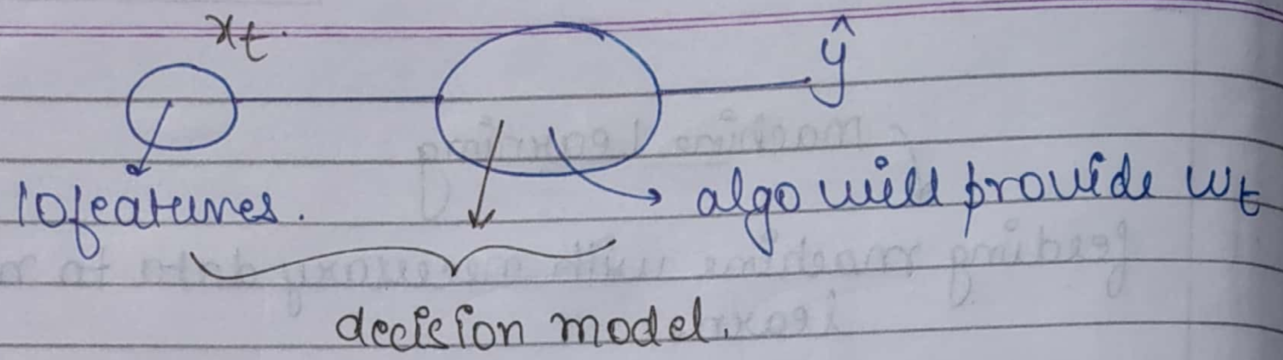
y → actual value

$\hat{y}$ → estimate

features →	x_1	x_2	- - - - - -	x_{10}	CL
1					A
2					A
3					B
⋮					A
⋮					B
200					B
					⋮

CL → (Class label)

200 → samples.



$$w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_{10} x_{10} = \hat{y}$$

where $w_0 + w_1 + \cdots w_{10} = 1$

Supervised learning

Classification/ pattern recogonition

- Class label is categorical.
- discrete values of class labels to be predicted.
- here we can differentiate objects into diff categories/ discrete values & not range of values.

eg:- Types of fishes (A, B, C, ...)

- y is an integral value

Regression/ estimate.

- used to predict continous scale
- continous value of class labels.

eg:- predicting price of a property based upon some features.

eg :- Weight (cont. variable).

- y is a real valued data

y → value to be predicted.

CLASSTIME Page No.
Date / /

NOTE :- Classifiers to be used in supervised learning are :-

(1) Naive Bayes classifier
(2) linear Reg.
(3) Logistic
(4) Neural Net.

- Unsupervised Learning :- (Clustering).

Here we aren't given labels but some useful patterns. (association)

grouping in some patterns

eg :- Basket - market example / Bread butter clustering.

→ Suppose an object has $x_1 \cdots x_n$ features.

$$X = x_1, x_2 \cdots x_n$$

and can be classified into k classes

$$C_1 C_2 \cdots C_k.$$

$P(C_i \mid X) \longrightarrow$ probability of class C_i for given values of x.

Using Bayes formula

$$\frac{P(X \mid C_i)\, P(C_i)}{P(X)} \longrightarrow \text{evidence}$$

for $i = 1, 2, 3, \cdots$

Denominator is fixed.
and numerator varies.

$$\boxed{P(C_i \mid X) \propto P(X \mid C_i)\, P(C_i)}$$

Posterior ↓ Likelihood ↓ → Priori

Find ↓

Let
$0.6 = P(C = C_1 \mid \langle x_1, x_2 \rangle)$
$0.3 = P(C = C_2 \mid \langle x_1, x_2 \rangle)$
$0.1 = P(C = C_3 \mid \langle x_1, x_2 \rangle)$

Sum = 1

max. argument

$$\hat{y} = \text{max. arg } P(C_i \mid X)$$

for $P(C_i)$ → If nothing is given, probability is distributed equally among all C_i's.

priori

(Derivation not imp.)

<u>or</u>

Calculate it from $\underbrace{\text{training}}_{\text{Sample}}$ data

1000 → A (400/1000)
1000 → B (500/1000)
1000 → C (100/1000)

$X = x_1, x_2 \cdots x_n$

$P(X|c_i) = P(<x_1 x_2 x_3 \cdots x_n>|c_i)$

$P(X|c_i)\,P(c_i) = P(X \cap c_i)$

$= P(\underbrace{x_1}_{\text{breaking}}\ \underbrace{x_2 \cdots x_n c_i}_{\text{every variable in intersection}})$

$P(x_1 | x_2 \cdots x_n c_i)\, P(x_2 \cdots x_n c_i)$

$= P(x_1 | x_2 \cdots x_n c_i)\, P(x_2 | x_3 \cdots x_n c_i)$
$P(x_3 \cdots x_n c_i)$

let $X = x_1 . x_2 . x_3$

$= P(x_1 | x_2 x_3 c_i)\, P(x_2 | x_3 c_i)\, P(x_3 c_i)$

$= P(x_1 | x_2 x_3 c_i)\, P(x_2 | x_3 c_i) . P(x_3 | c_i)\, P(c_i)$ — ①.

NOTE:- Naive Bayes classification applies on those attributes that are independent of each other.

(Length and weight are not independent)
and increase in length increases weight.

$$P(A \cap B) = \underbrace{P(A|B) \cdot P(B)}$$

Given - If A and B are independent events then $P(A \cap B) = P(A) \cdot P(B)$.

eqn (1) can be written as :-

$$= P(x_1|ci) \cdot P(x_2|ci)\, P(x_3|ci)\, P(ci)$$

(as x_1 is independent of x_2 and x_3) {for Naive's Bayes classification}.

$$= \prod_{j=1}^{n} (P(x_j|ci))\, P(ci)$$

Imp:

$$\boxed{V_{NB} = \text{argmax}\, P(V_j) \prod_i P(a_i/V_j)}$$

(Here $V_j = c_i$ and $a_i = x_j$)

Example

	x_1	x_2	x_3	x_4	C.L (y)	
Day	Outlook	Temp	Humidity	Wind	Play Tennis	
D1	Sunny	Hot	High	Weak	No	
D2	Sunny	Hot	High	Strong	No	
D3	Overcast	Hot	High	Weak	Yes	1
D4	Rain	Mild	High	Weak	Yes	2
D5	Rain	Cool	Normal	Weak	Yes	3
D6	Rain	cool	Normal	Strong	No	

D7	Overcast	Cool	Normal	Strong	Yes	4
D8	Sunny	Mild	High	Weak	No	
D9	Sunny	Cool	Normal	Weak	Yes	5
D10	Rain	Mild	Normal	Weak	Yes	6
D11	Sunny	Mild	Normal	Strong	Yes	7
D12	Overcast	Mild	High	Strong	Yes	8
D13	Overcast	Hot	Normal	Weak	Yes	9
D14	Rain	Mild	High	Strong	NO	

$V_{NB} = \underset{V_j \in \{Yes, No\}}{\arg\max} \longrightarrow P(V_j) \prod_i P(a_i | V_j)$

Here from Training Data,

$P(Yes) = \frac{9}{14}$ ⇐ max.

$P(NO) = 5/14.$

Given → (Outlook = Sunny
Humidity = High
Wind = Strong
Temp = Cool)

→ Decision model will have 22 parameters irrespective of any size of data.

P(Yes), P(No), P(Sunny | Yes). P(Sunny|No)
P(Overcast | Yes). P(Overcast| No)
P(Rain | Yes). P(Rain | No)

This way we will have
$2 + 6 + 6 + 4 + 4 = \underline{22}$.

CLASSTIME	Page No.
Date / /	

$P(Yes) = \frac{9}{14}$ $\quad$ $P(No) = \frac{5}{14}$

$P(outlook = sunny | Yes) = \frac{2}{9}$

$P(Sunny | No) = \frac{3}{5} \quad 5 \longrightarrow (14-9)$

$P(Outlook = Sunny) = \frac{5}{14}$

$P(Yes | Outlook = Sunny) = \frac{2}{5}$

$$\frac{P(Sunny|Yes) \cdot P(Yes)}{P(Sunny)} = \frac{\frac{2}{9} \times \frac{9}{14}}{5/14} = \frac{2}{5}$$

$P(Temp = cool | Yes) = \frac{3}{9}$

Total (Yes = 9)

$P(cool | No) = \frac{1}{5}$

$P(Wind = Strong | Yes) = \frac{3}{9}$

$P(Wind | No) = \frac{3}{5}$

$P(High | Yes) = \frac{3}{9}$ $\quad$ $P(High | No) = \frac{4}{5}$.

(Yes)

$P(Yes | X') = \frac{9}{14} \times \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} = \frac{2}{378}$

proportion (α) = 0.005291

$P(No | X') = \frac{5}{14} \times \frac{3}{5} \times \frac{1}{5} \times \frac{3}{5} \times \frac{4}{5} = \frac{18}{875}$

(no)

Proportion (α) = 0.0205

CLASSTIME Page No.
Date / /

Final answer]

$$V_{NB} = \text{max of } (\text{Yes or no})$$

$$= \underline{\text{No}}$$

$\boxed{\text{class label} = \text{No}}$

Probability of No and Yes

$$P(X) = \sum_{i=1}^{n} P(X|C_i)\, P(C_i)$$

$$P(C_k \mid X) = \frac{P(X|C_k) \cdot P(C_k)}{P(X)}$$

$$P(\text{Yes} \mid X) = \frac{0.00529}{0.00529 + 0.0205} = \underline{0.205}$$

$$P(\text{No} \mid X) = \frac{0.0205}{0.00529 + 0.0205} = \underline{0.794}$$

4 Jan. 19

CLASSTIME Page No.
Date / /

Practical

> help. start ()

1to 5

> a = 1:5 (1D array of length 5)
> a[1] = 7

> a = seq (from = 1, to = 10, by = 2 {1 3 5 7 9}
> a = seq (1, 10, 2)

> ? seq (Help) // ? name

> a // print
[1] 7 2 3 4 5.

→ If we don't know exact function.
?? name ⇒ displays everything related to this.

To know about fname's declaration and definition.

> fnname
{ outputs full function}.

- For initializing 1D array / column vector.

> a = c (20, 25, 24, 23) { no need to define datatype but only one type of array }
> a
[1] 20 25 24 23

- For sequential array {1D array}

CLASSTIME Page No.
Date / /

① > a = seq (from=1, to=10, by 2)
> a
[1] 1 3 5 7 9

or

② > a = 1:5 (1 to 5)
> a
[1] 1 2 3 4 5

> a = 1:100
> a
[1] 1 2 3 4 5 ------ 100.

> b = c ("Everyone", "likes", "machines")
> b
[1] "Everyone" "likes" "machines"

If > b = c (2, "Everyone")
> b
[1] "2" "Everyone".

> example ("seq") // displays usage of seq.

> help (seq) // tells about sequence & its methods.

> v = c (0, 1, 1, 2, 3, 5, 8, 13, 21, 34)
> x = v[3]
> x
[1] 1

> v[1:5]
[1] 0 1 1 2 3

// fetching sub array from index 1 to 5.

CLASSTIME Page No.
Date / /

```
> v [c(1,5,2,8)]        // extract elements of
   ↳                       these indexes.
     [1] 0 3 1 1 3

> v [-1]                // removes index value from
   ↳  [1] 1 1 2 3 5 8 13 21 34   v and doesnt
                                  make change in v.

> v [-1:-2]   // v [-(1:2)]  { remove index
   ↳ [1] 2 3 5 8 13 21 34       value from 1 to 2 }

> v [-c(1,3)]   // remove these particular
                          index values.
   ↳ [1] 3 18 13 21 24.

> v1 = c(1,2,4)
> v2 = c(34,56,2)
> v3 = c (v1, v2)      // combine v1 and v2.
> v3
      [1] 1 2 4 34 56 2

> c (1,2,3.1)          // automatic conversion
   [1] 1.0 2.0 3.1      to higher DT.

> v1 = c (1,2,3)       } comparing index
> v2 = c (1,2,4)         value of 1 with
> v1 = = v2              index value of 2
  [1] True True False    and then returning
                         true.

v1 = 1
v2 = "1"
v1 = = v2
          [1] True.
```

CLASSTIME Page No.
Date / /

⟹ Addition ↓
① If size is same for both arrays add corresponding index values.

② If size mismatches, then first add till same size and then start again adding from start.
and also display warning.

⟶ > v_1 = c(1,2,3)
> v_2 = c(1,2,4)
> $v_1 + v_2$
[1] 2 4 7.

> v_1 = c(1,2,3)
> v_2 = c(1,2,3,4)
> $v_1 + v_2$ ⟶ Warning message.
[1] 2 4 6 5 → 4+1

> a
[1] 1 2 3.
> b
[1] 3 4 5
> a * b
[1] 3 8 15
(1X3) (2X4) (5X3)

multiply & divide in similar manner

$\left(\frac{5}{3}\right)$

> a / b

[1] 0·33 0·500 0·6000

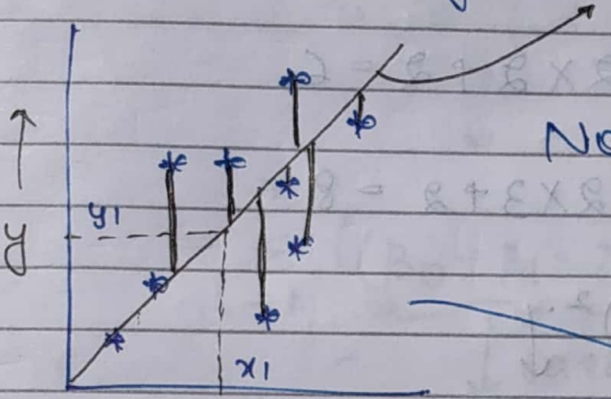
CLASSTIME Page No:
Date / /

11 Jan, 19

* Linear Regression -

Forming a line given the training data, we need to know slope and intercept

$$y = mx + c$$



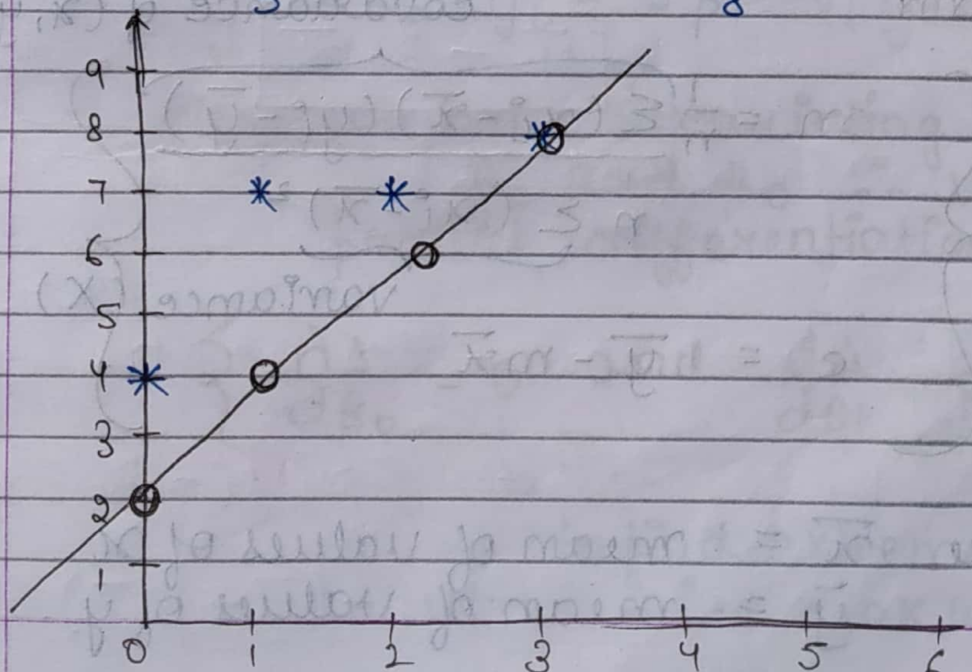
Now if some x_1 comes, we can estimate a price of y. denoted by y_1.

line must be such that Sum of vertical distances must be least.

Ex -

	I/P	O/P
	x	y
To predict	0	4
the best	1	7
Suitable	2	7
line	3	8

↳



Assuming $m = 2$ and $c = 2$

$y = mx + c$

(estimated val.) $\hat{y}$

$(0,4)$ $\hat{y} = 2 \times 0 + 2 = 2$

$(1,7)$ $\hat{y} = 2 \times 1 + 2 = 4$

$(2,7)$ $\hat{y} = 2 \times 2 + 2 = 6$

$(3,8)$ $\hat{y} = 2 \times 3 + 2 = 8$

Now $(\hat{y} - y)^2$

we can either take $| \, |$ or do $(\,)^2$ of errors.

$$\left.\begin{matrix} 4 \\ 9 \\ 1 \\ 0 \end{matrix}\right\} \text{adding them} = \underline{14}.$$

$$\boxed{\Delta = \sum_{i=1}^{n} (\hat{y_i} - y_i)^2}$$

error

$$\left\{ \begin{aligned} m &= \frac{\overbrace{\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})}^{\text{covariance of } (x,y)}}{\underbrace{\frac{1}{n} \sum (x_i - \bar{x})^2}_{\text{variance } (X)}} \\ c &= \bar{y} - m\bar{x} \end{aligned} \right\}$$

where $\bar{x}$ = mean of values of x

$\bar{y}$ = mean of values of y.

CLASSTIME Page No:
Date / /

Derivation-

$$\Delta = \sum_{i=1}^{n} (\hat{y_i} - y_i)^2$$

Considering $y = mx + c$
Put $c = \beta_0$
and $m = \beta_1$

$$\therefore \Delta = \sum_{i=1}^{n} ((\beta_0 + \beta_1 x_i) - y_i)^2$$

$$= \sum_{i=1}^{n} (\underbrace{(\beta_0 + \beta_1 x_i)^2}_{(a+b)^2} + y_i^2 - 2y_i(\beta_0 + \beta_1 x_i))$$

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

$$= \sum_{i=1}^{n} \beta_0^2 + \beta_1^2 \sum_{i=1}^{n} x_i^2 + 2\beta_0\beta_1 \Sigma x_i + \Sigma y_i^2 - 2\beta_0 \Sigma y_i - 2\beta_1 \Sigma x_i y_i.$$

Here we have 2 variating quantities β_0 and β_0 $\therefore$ we perform partial differentiation

$$\left\{ \frac{\partial \Delta}{\partial \beta_0} = 0 \text{ and } \frac{\partial \Delta}{\partial \beta_1} = 0 \right\}$$

to find extreme pts.
i.e max or min. pt.

CLASSTIME Page No.
Date / /

Diff. wrt β_0

$$\frac{\partial \Delta}{\partial \beta_0} = 2\sum_{i=1}^{n} \beta_0 + 2\beta_1 \Sigma x_i - 2\Sigma y_i$$

(diff. wrt β_1) $$\frac{\partial \Delta}{\partial \beta_1} = 2\sum_{i=1}^{n} \beta_1 x_i^2 + 2\beta_0 \Sigma x_i - 2.\Sigma x_i y_i$$

Now $\frac{\partial \Delta}{\partial \beta_0} = 0$

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

$$n\beta_0 + \beta_1 \Sigma x_i - \Sigma y_i = 0$$

$$\beta_0 = \frac{\Sigma y_i - \beta_1 \Sigma x_i}{n}$$

$$= \frac{\Sigma y_i}{n} - \frac{\beta_1 \Sigma x_i}{n}$$

(mean of y)

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

Now $\frac{\partial \Delta}{\partial \beta_1} = 0$

$$\beta_1 \sum_{i=1}^{n} x_i^2 + \beta_0 \Sigma x_i - \Sigma x_i y_i = 0.$$

$$\beta_1 \sum_{i=1}^{n} x_i^2 = \Sigma x_i y_i - \beta_0 \Sigma x_i$$

$$= \Sigma x_i y_i - (\bar{y} - \beta_1 \bar{x}) \Sigma x_i$$

$$\beta_1 \Sigma x_i^2 + \left(\frac{\Sigma y_i}{n} - \beta_1 \frac{\Sigma x_i}{n}\right)\Sigma x_i - \Sigma x_i y_i = 0$$

$$\beta_1\left(\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right) + \frac{\Sigma x_i \Sigma y_i}{n} - \Sigma x_i y_i = 0.$$

$$\beta_1 = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}$$

(Slope=m)

$$\boxed{= \frac{n \Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n \Sigma x_i^2 - (\Sigma x_i)^2}}$$

Now $\beta_1 = \dfrac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$

$n\left(\frac{\Sigma x_i}{n} \cdot \frac{\Sigma y_i}{n}\right)$

$$\frac{\Sigma x_i y_i - \Sigma x_i \bar{y} - \Sigma \bar{x} y_i + \Sigma \bar{x}\bar{y}}{\Sigma x_i^2 + \Sigma \bar{x}^2 - 2\Sigma x_i \bar{x}}.$$

$$\frac{\Sigma x_i y_i - \cancel{\frac{\Sigma x_i \Sigma y_i}{n}} - \frac{\Sigma x_i}{n} y_i + \cancel{\frac{\Sigma x_i \Sigma y_i}{n}}}{\Sigma x_i^2 + \Sigma \frac{x_i^2}{n^2} - 2 \frac{\Sigma x_i \Sigma x_i}{n}}$$

n

$$\cancel{= \frac{\Sigma x_i y_i\left(1 - \frac{1}{n}\right) + \Sigma x_i \Sigma y_i \left[\frac{1}{n} \ldots\right]}{\Sigma x_i^2\left(1 + \frac{1}{n}\right) - 2\frac{(\Sigma x_i)^2}{n}}}$$

$$= \frac{\Sigma x_i y_i\left(1 - \frac{1}{n}\right)}{\Sigma x_i^2 - \frac{1}{n}(\Sigma x_i)^2}$$

CLASSTIME Page No.
Date / /

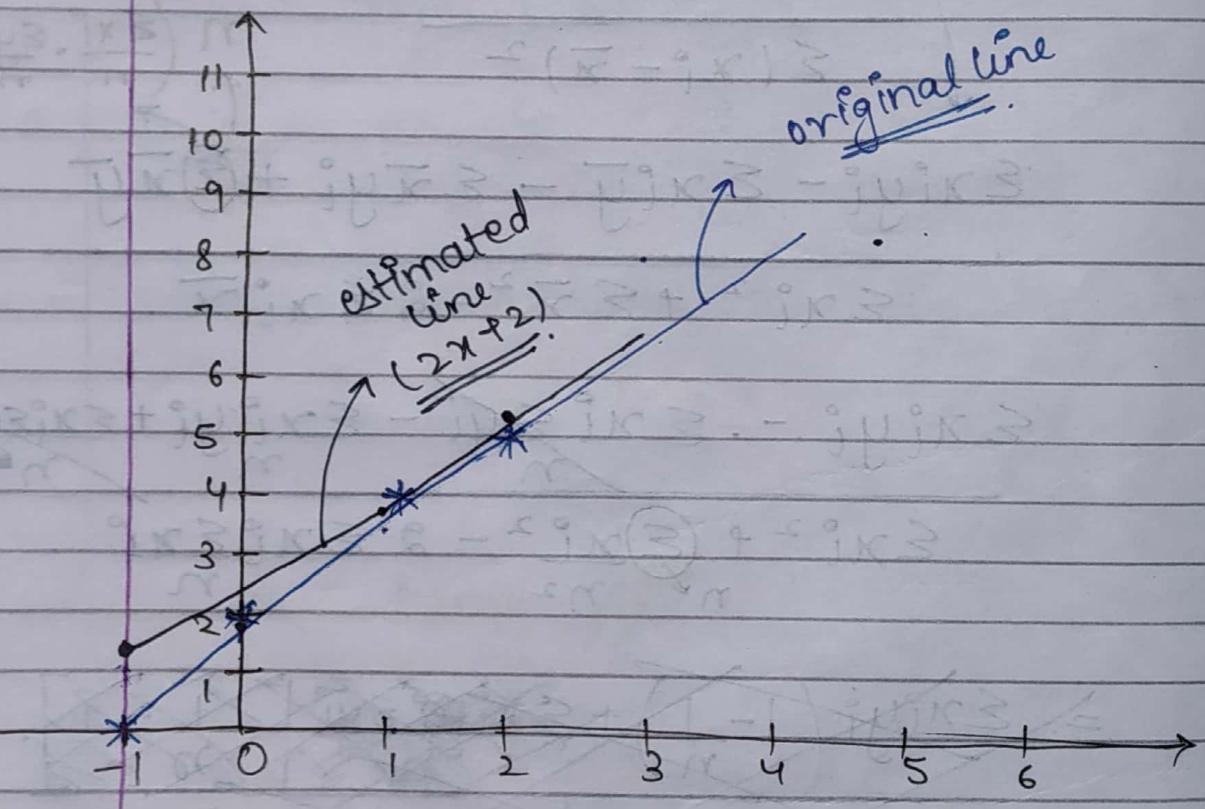
$$\boxed{= \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}}$$

Q. (-1, 0), (0, 2), (1, 4), (2, 5)

↓

Find the least regression line

x	y
-1	0
0	2
1	4
2	5



Choosing $m = 2$ and $c = 2$

$$\hat{y} = 2x + 2$$

CLASSTIME Page No.
Date / /

x	y	$x_i-\bar{x}$	$y_i-\bar{y}$	$\Sigma(x_i-\bar{x})(y_i-\bar{y})$
-1	0	-3/2	-11/4	33/8
0	2	-1/2	-3/4	3/8
1	4	1/2	5/4	5/8
2	5	3/2	9/4	27/8

$\bar{x} = \frac{1}{2}$, $\bar{y} = \frac{11}{4}$

$\therefore \frac{33+3+5+27}{8} = \frac{68}{8}$

$\Sigma(x_i - \bar{x})^2$

$$\left.\begin{matrix} 9/4 \\ 1/4 \\ 1/4 \\ 9/4 \end{matrix}\right\} \frac{20}{4} = 5.$$

$$\therefore m = \frac{\Sigma(x_i - x)(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{68}{8 \times 5} = \frac{17}{10} = 1.7.$$

$$c = \bar{y} - m\bar{x}$$

$$= \frac{11}{4} - \frac{17}{10}\left(\frac{1}{2}\right) = \frac{38}{20} = 1.9.$$

Now $\hat{y} = 2x + 2$

$\therefore$ Plotting line, we get

x	$\hat{y}$
-1	1.2
1	3.6
0	1.9
2	5.3.

For predicting value of y for $x = 3$

$\hat{y} = 2(3) + 2$
$= 8.$

{Vector - 1D array}
{Matrix - 2D array}

a = c (2, 3, 5, 56, 7, 4)
↓
vector

→ $a = \text{matrix}(c(1,2,3,4), \underbrace{2,2}_{m \times n})$

$$\begin{array}{c|cc} & [\,,1] & [\,,2] \\ \hline [1,\,] & 1 & 3 \\ [2,\,] & 2 & 4 \end{array}$$

a = matrix (1:12 , 2, 6)

Starts filling rowwise

$$\begin{array}{c|cccccc} & [\,,1] & [\,,2] & - & - & - & [\,,6] \\ \hline [1,\,] & 1 & 3 & 5 & 7 & 9 & 11 \\ [2,\,] & 2 & 4 & 6 & 8 & 10 & 12 \end{array}$$

→ a = 1:12

$\underbrace{\text{dim(a)} = c\,(2, 2, 3)}_{\text{3 dimensions}}$

$$\underbrace{\underset{,\,,1}{\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}} \quad \underset{,\,,2}{\begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix}} \quad \underset{,\,,3}{\begin{bmatrix} 9 & 11 \\ 10 & 12 \end{bmatrix}}}_{\text{3 matrix of 2X2}}$$