Linear Regression

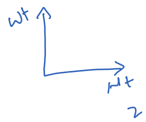
Linear / Regression → o/p is continuous

$y = mx + c$

Assumption

Feature are linearly varying with label

$x \uparrow, y \uparrow$
$x \uparrow, y \downarrow$ } Linear



E.g, * Ht v/s Wt
Ht↑ => Wt↑

* Sqft v/s Price
sqft↑ => price

* No. of years v/s Price
No. of years ↑ , Price↓

Linear ⇐ $y = 5x + 2$ → Order = 1
Quadratic ⇐ $y = 3x^2 - 8$ → order = 2
cubic ⇐ $y = x^3 + 5$ → Order = 3
⋮
$y = 3x^3 + 2x^2 + 5x - 2$
↳ order = 3

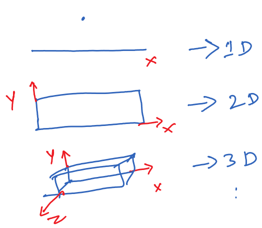
$y = 2x + 5$

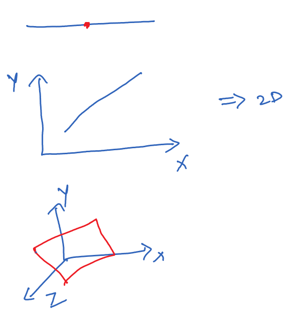
$x = 1 \Rightarrow y = 2(1) + 5 = 7$
$x = 2 \Rightarrow y = 2 \times 2 + 5 = 9$
$x = 3 \Rightarrow y = 2 \times 3 + 5 = 11$

$y = x^2 \Rightarrow$ Eqⁿ of parabola.
$y = 1$
$y = 4$
$y = 9$

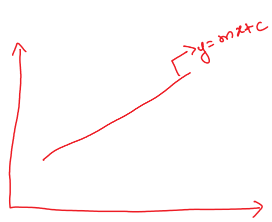
$x \uparrow, y \uparrow$

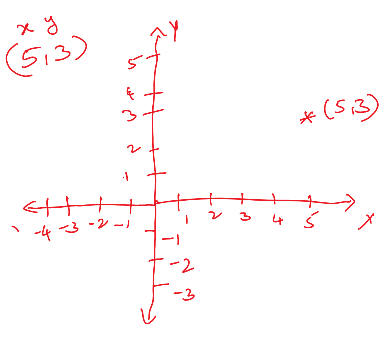
2D → 1D
3D → 2D





$y = mx + c$





$y = mx + c$

m ↑ Co-efficient of x / slope

c ↑ y-intercept

$m = \tan\theta$

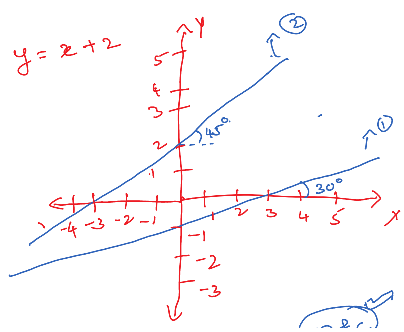
$m_1 = \tan 30$
$= 0.57$

$m_2 = \tan 45$
$m_2 = 1$

$c_1 = -1$

$c_2 = 2$

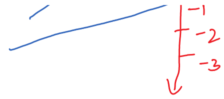
$y = x + 2$



$= 0.5$, $\quad m_2 = 1$

$c_1 = -1 \quad c_2 = 2$

$y = 0.57x - 1 \quad y = 1x + 2$



$\boxed{m \, c}$

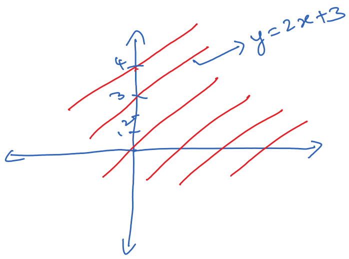
$y = 2x + \boxed{3}$

$m = 2 \quad c = 3$

$\tan\theta = 2$

$\theta = \tan^{-1} 2$

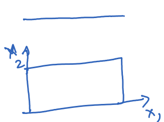
$\theta = 63.43$



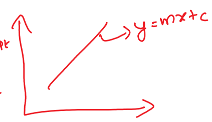
Line ← 1D ← $y = mx + c$

Plane ← 2D ← $y = m_1x_1 + m_2x_2 + c$

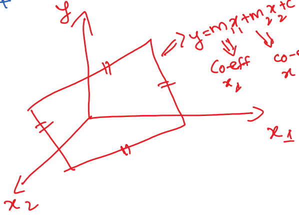
Hyperplane $\begin{cases} y = m_1x_1 + m_2x_2 + m_3x_3 + c \\ y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + c \\ \vdots \end{cases}$



$ax^2 + bx + c$



$y = m_1x_1 + m_2x_2 + c$ — m_1: Co-eff x_1, m_2: co-eff x_2



4D ⇒ 3D
5D ⇒ 4D

Σ

$y = m_1x_1 + m_2x_2 + m_3x_3 + \dots m_nx_n + c$

$y = \sum_{i=1}^{n} m_ix_i + c$

$n = 3 \Rightarrow y = \sum_{i=1}^{3} m_ix_i + c$

$= 3 \Rightarrow y = m_1x_1 + m_2x_2 + m_3x_3 + c$

ML

$A = [\quad]_{m \times n} \quad B = [\quad]_{n \times p}$

$m \times \boxed{n \quad n} \times q$

$y = m_1x_1 + m_2x_2 + m_3x_3 + \dots m_nx_n + c$

$m = [m_1 \; m_2 \dots m_n]_{1 \times n} \qquad 1 \times \boxed{n \quad 1} \times n$

$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}_{1 \times 3} \quad \begin{bmatrix} 2 & 1 \\ 3 & 2 \\ 4 & 3 \end{bmatrix}_{3 \times 2}$

$1 \times \boxed{3 \quad 3} \times 2$

1×2

$x = [x_1 \; x_2 \dots x_n]_{1 \times n}$

$m^T = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix}_{n \times 1} \qquad m^T \times x \quad n \times \boxed{1 \quad 1} \times n$

$m^Tx = m_1x_1 + m_2x_2 + m_3x_3 + \dots + m_nx_n$

$y = m^Tx + c$

$m = w, \quad c = b$

$\boxed{y = w^Tx + b}$ → General eq^n of Hyperplane

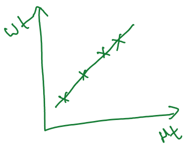
$y = wx + b$ → Eq^n of line.

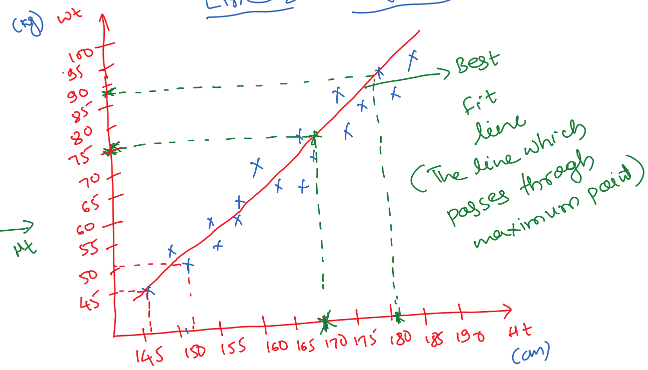
$y = w_1x_1 + w_2x_2 + b$ → " " plane

$y = w_1x_1 + w_2x_2 + w_3x_3 + b$ →

Linear Regression

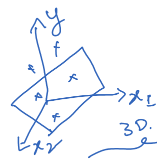
2D →
3D →
4D →





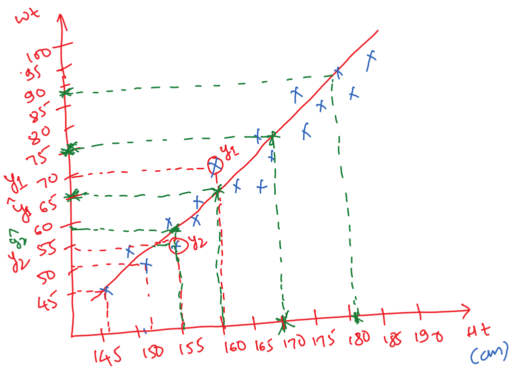
I/P Ht	O/P wt
145	45
150	52
⋮	⋮
170	? ↳ 75 kg.
182	? ↳ 82 kg

$y = mx + c$ →



3D.

x_1	x_2	y



Actual Ht → 160 cm
y_1 = Actual wt → 70 Kg
$\hat{y}_1$ = predicted wt → 65 Kg

Error = Actual − Predicted
= 70 − 65
= 5

Error = $y - \hat{y}$

Error = $y_2 - \hat{y}_2$
= 55 − 59
= −4

Error = $y - \hat{y}$

Cost function = $\sum (y - \hat{y})^2$

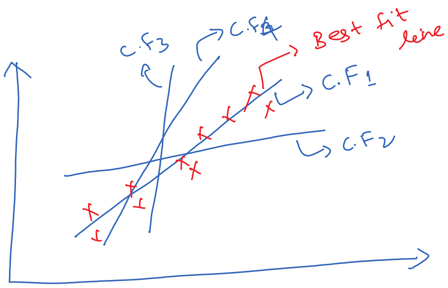
$(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + \ldots$

= 5 + 4 + 3 + 2 = 14

5 − 4 = 1

5 + 4 = 9

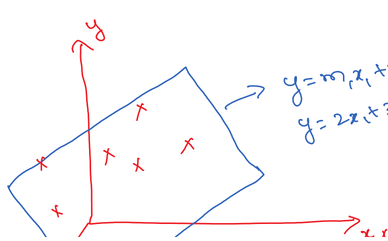
25 + 16 = 41



Square ⇒ Value change

Moto is to choose best fit line ⇒.

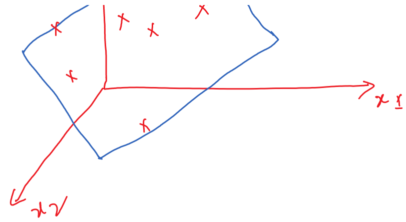
Choose that line which is giving least Cost function (Error)



→ $y = m_1x_1 + m_2x_2 + c$

$y = 2x_1 + 3x_2 - 8$

I/P x_1	I/P x_2	O/P y
—	—	—
—	—	—
—	—	—
—	—	—
—	—	—
5	3	? ⇒ 11



–	–	–
–	–	–
–	–	–
5	3	? => 11

$y = 2\times5 + 3\times3 - 8$
$= 10 + 9 - 8$
$= 11$

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

$|y-\hat{y}|$

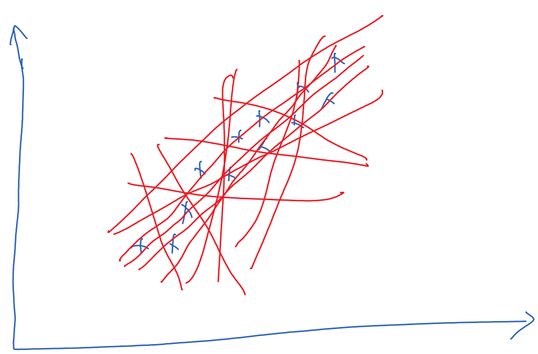
$\frac{\Sigma |y-\hat{y}|}{n}$

$4 \Rightarrow \boxed{16}$ ... $\overline{\quad}\; n$

Cost function

1. Square Error. $= \Sigma (y-\hat{y})^2$
2. Mean square Error (MSE) $= \frac{\Sigma (y-\hat{y})^2}{n}$
3. Root Mean square Error (RMSE) $= \sqrt{MSE} = \sqrt{\frac{\Sigma (y-\hat{y})^2}{n}}$
4. Mean Absolute deviation (MAD) $= \frac{\Sigma |y-\hat{y}|}{n}$

* Choose that line which is giving least MSE.



Infinite number of lines

Gradient Descent

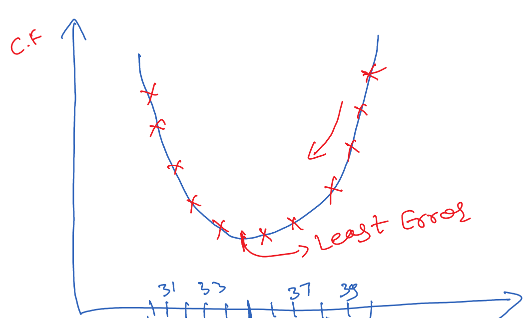
Random guess

Step by step

40.	30
39	31
38	32
37	33
36	34
$\boxed{35}$	$\boxed{35}$

40 $\xrightarrow{-10}$ 30 $\xrightarrow{+2}$ 32 $\xrightarrow{+6}$ 38 $\xrightarrow{-5}$ 33 → 37 → 34 → 36 → $\boxed{35}$

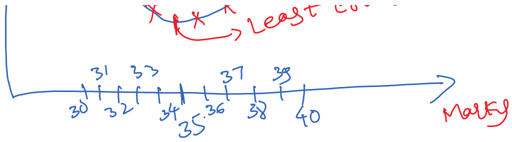
40 − 35 = 5
39 − 35 = 4



New = old − step size
New = 40 − 1 = 39

New = old − stepsize
= 39 − 1

New = old − stepsize
= 30 − (−1)



New = old − stepsize
= 30 − (−1)
= 31

O/p → $y = mx + c$ ← I/p

m → slope / co-eff. of x / Gradient

c → y-intercept

x_1	x_2	x_3	y

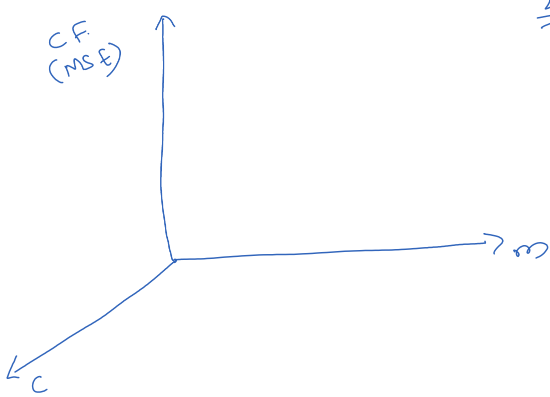
x I/p, y O/p

$y = m_1 x_1 + m_2 x_2 + m_3 x_3 + c$

fixed → $y = mx + c$ ← fixed

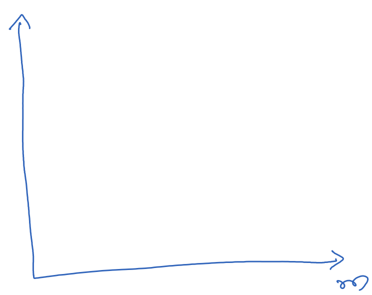
m, c → Which are giving least MSE then I get a line which gives least MSE

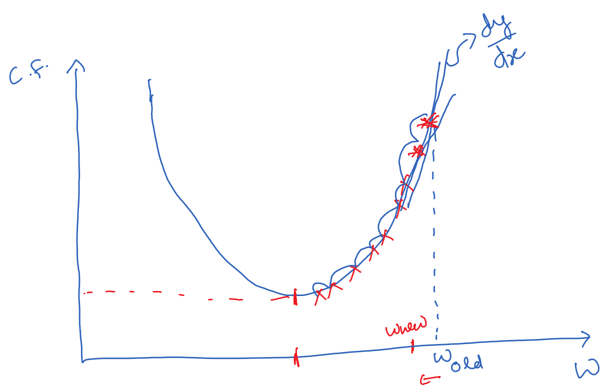
3D



⟹

2D



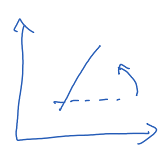


slope, $m = \tan\theta$

$\frac{dy}{dx} = \frac{d(CF)}{dw_{old}}$

$m = \text{slope} = \tan\theta = \frac{dy}{dx}$

0 → 90°
↳ +ve
$m = \tan\theta$
$m = 0$
$m = 1$
$m = \infty$

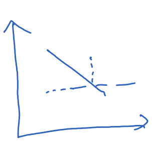


$w_{new} = w_{old} - \text{Step size}$
$= w_{old} - \eta \frac{d(CF)}{dw_{old}}$

η → Learning Rate
$= 0.1, 0.01$

$w_{new} = w_{old} - (0.1)(+ve)$

$w_{new} < w_{old}$



$>90°$
slope will be (−ve)
$m = \tan 90$

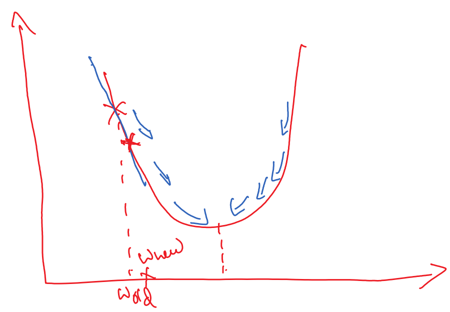
Wold

$w_{new} = w_{old} - (0.1)(+ve)$

$w_{new} < w_{old}$

$w_{new} \approx w_{old} \Rightarrow$ Stop the process

$w \Rightarrow$



$w_{new} = w_{old} - \eta \frac{dL}{dw_{old}}$

$= w_{old} - (0.1)(-ve)$

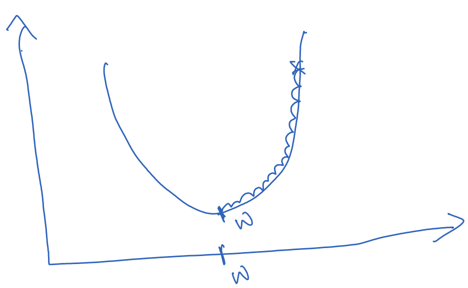
$= w_{old} + \boxed{}$

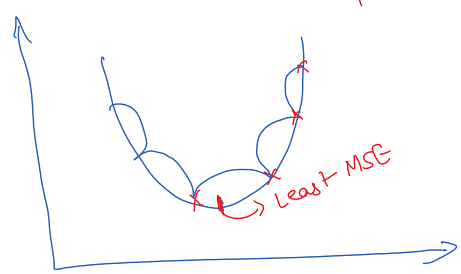
$w_{new} > w_{old}$

Stepsize

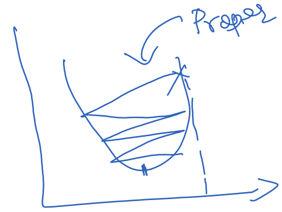
$0.1 \times 50 = 5$

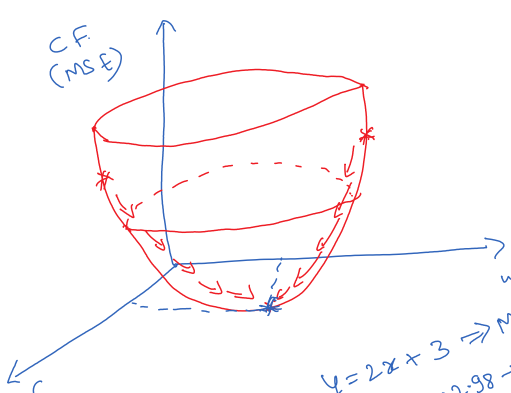
Learning Rate





Not Proper method





$w_{new} = w_{old} - \eta \frac{d(C.F)}{dw_{old}}$

$c_{new} = c_{old} - \eta \frac{d(C.F)}{dc_{old}}$

w , c

$w = [2, 1.98 \quad]$

$c = [3, 2.98 \quad]$

$y = 2x + 3 \Rightarrow MSE = CF$

$y = 1.98x + 2.98 \rightarrow MSE = CF$

m, c

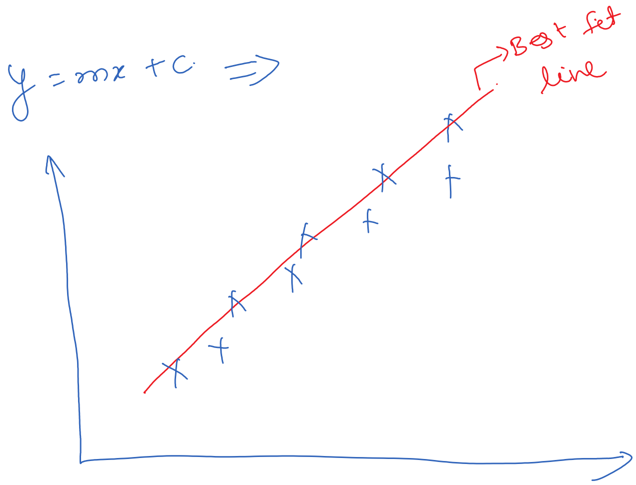
m , c.
2 3

$m_{new} \approx m_{old}$

$y = 1.5x + 2 \Rightarrow$ least MSE

$\Rightarrow$ Best fit

$y = mx + c \Rightarrow$



$y = mx + c$
$= 2x + 3$

$m = 2$
$\tan\theta = 2$
$\boxed{\theta = 62^\circ}$

Pipeline for Project

* Understand Business Case & do domain Analysis
* EDA (Exploratory data Analysis)
  - → statistical (Mean, std, percentile)
  - → Visualization
* Data Preprocessing
  - → Checks for null values & corrupted data
  - → Converting categorical data into Numerical data
  - → Handling with outliers
  - → Scaling (If required)
  - → Balancing (If required) (Classification pr. only)
* Model Creation
* Model Evaluation
* Hyperparameter Tuning
* Deploy the model

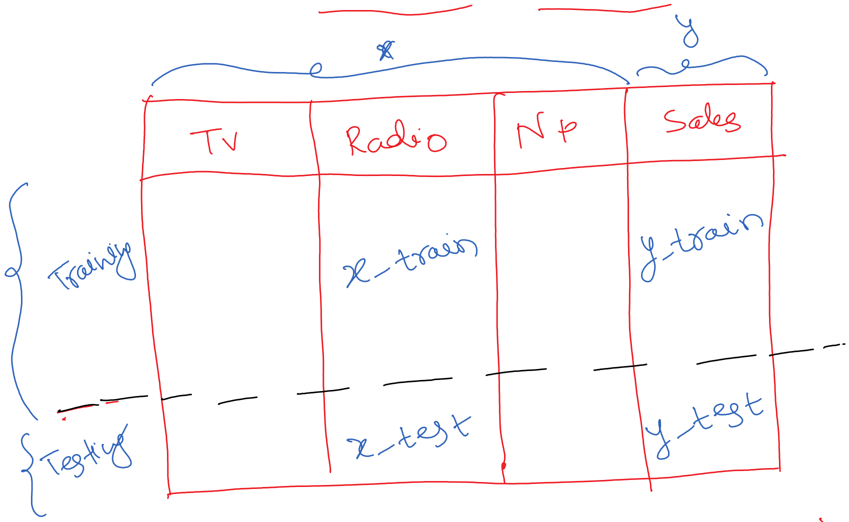
Model Creation
Model Evaluation

Model Creation

200 → 50
↓
150
75 : 25
200

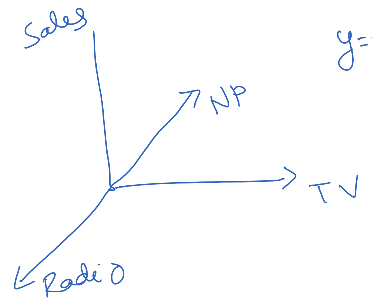
Model Creation

Model Evalution



1 column = 1 D

Wt
Ht
$y = m \times Ht + c$



x_train & y_train →

$y = m_1 \times TV + m_2 \times R + m_3 \times NP + C$

x_test ⇒ y_pred

Compare → y_test

TV = 200 , R = 150 NP = 100 ⇒ $y = 200m_1 + 150m_2 + 100m_3 + c$

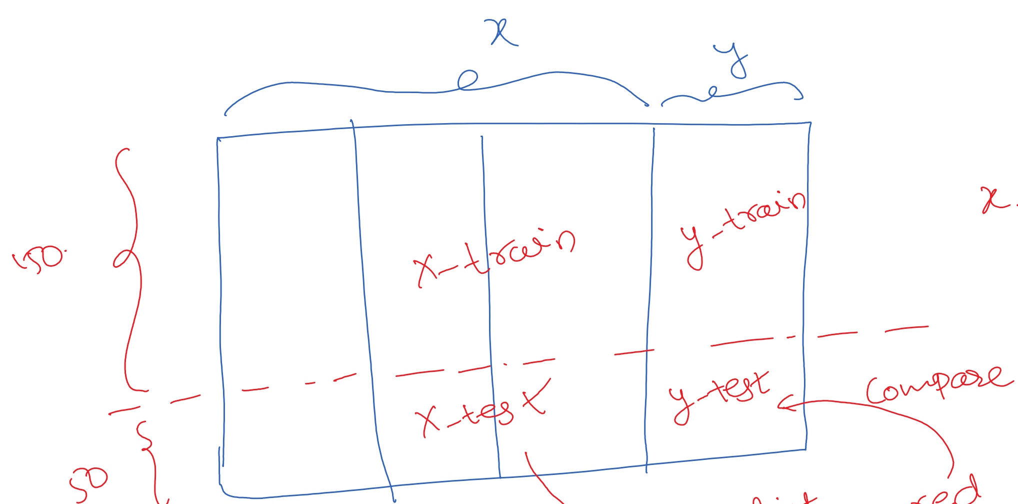
Predicted ⇐ y_pred = 20

Actual ⇐ y_test = 22

Error = Actual - Predicted

$\boxed{x \,\&\, y}$

200 → 50
↓
150



x_train & y_train.

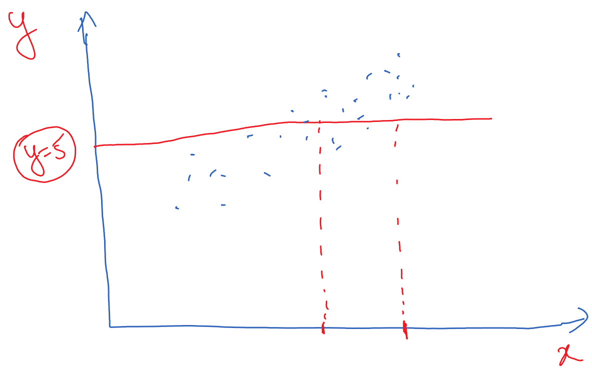
Predict x_test = y_pred

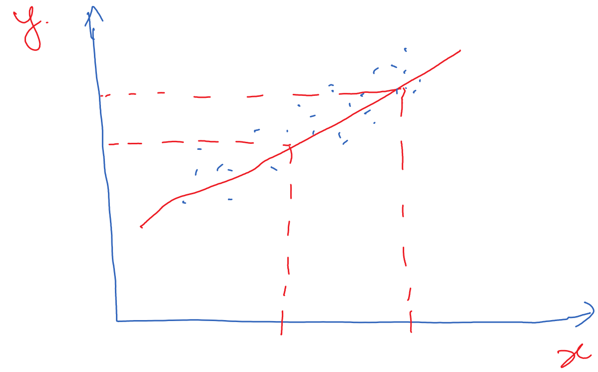
x	y
Q	A

x: $\boxed{Q}$ ⇒ y_pred: $\boxed{A}$ y_test

R^2 – Statistics

Average Model / Dumb Model





Residue => Error

y

$\hat{y}$

$\bar{y}$

$$R^2 = 1 - \frac{\text{Residue by your model}}{\text{Residue by dumb model}}$$

$$= 1 - \frac{\Sigma(y-\hat{y})^2}{\Sigma(y-\bar{y})^2}$$

Case 1: Res. by your model $>$ Res. by dumb. model.

$$R^2 = 1 - \frac{\Sigma(y-\hat{y})^2 \uparrow}{\Sigma(y-\bar{y})^2 \downarrow} \Rightarrow > 1$$

$R^2 < 0$ (-ve)

Case 2: Res. by your model $<$ Res. by dumb. model.

$$R^2 = 1 - \frac{\Sigma(y-\hat{y})^2 \downarrow}{\Sigma(y-\bar{y})^2 \uparrow} \Rightarrow < 1.$$

$R^2 > 0$ (+ve)

Case 3: Res. by your model $=$ Res. by dumb. model.

R^2

Case 3: Res. by your model = dumb. model.

$$R^2 = 1 - \frac{\cancel{\sum (y-\hat{y})^2}}{\cancel{\sum (y-\bar{y})^2}}$$

$$R^2 = 0$$

Case 4: Res by your model = 0

$$R^2 = 1 - \frac{\cancel{\sum (y-\hat{y})^2} \rightarrow 0}{\sum (y-\bar{y})^2}$$

$$R^2 = 1$$

$R^2 < 0 \rightarrow$ Worst than dumb model

$R^2 = 0 \rightarrow$ Same as dumb model

$R^2 > 0 \rightarrow$ Better than dumb model

$R^2 = 1 \rightarrow$ Res. by your model is zero

$R^2 \in (0, 1)$ → R^2 should be nearer to 1

$\boxed{0.9}$ $\boxed{0.85}$ $\boxed{0.95}$

$\boxed{0.2}$

$$R^2 = 1 - \frac{\sum (y-\hat{y})^2}{\sum (y-\bar{y})^2}$$

$\hat{y} = m_1x_1 + m_2x_2 + \dots + m_nx_n + c$

(column)

$\hat{y} = m_1x_1 + m_2x_2 + m_3x_3 + c$

$\hat{y} = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + c$

$$\frac{\sum (y-\hat{y})^2}{\sum (y-\bar{y})^2}$$

$5 - 3 = 2$

$5 - 4 = 1$

$\frac{3}{5}$ $\frac{2}{5}$

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

$\hat{y}\uparrow \Rightarrow (y-\hat{y})\downarrow \Rightarrow \frac{(y-\hat{y})^2 \downarrow}{(y-\bar{y})^2} \downarrow$

$\frac{3}{5}$ $\frac{2}{5}$

$= 0.6$ 0.4

$\Downarrow$

$1 - \frac{(y-\hat{y})^2}{(y-\bar{y})^2} \downarrow \uparrow$

R^2-score increasing. $\Rightarrow$ R^2-score is unstable.

Adjusted R^2-score $= 1 - \frac{(1-R^2)(n-1)}{(n-p-1)}$

$R^2 = R^2$-score

$n =$ No. of data pts in testing

$p =$ No. of features