AI:

Eliza

Rule Based,
AI on if-else
IBM Watson

AIML: Artificial Intelligence Markup language

AI

1- Machine Learning

↓

Data

↓

Pre-processing

↓

Training

↓

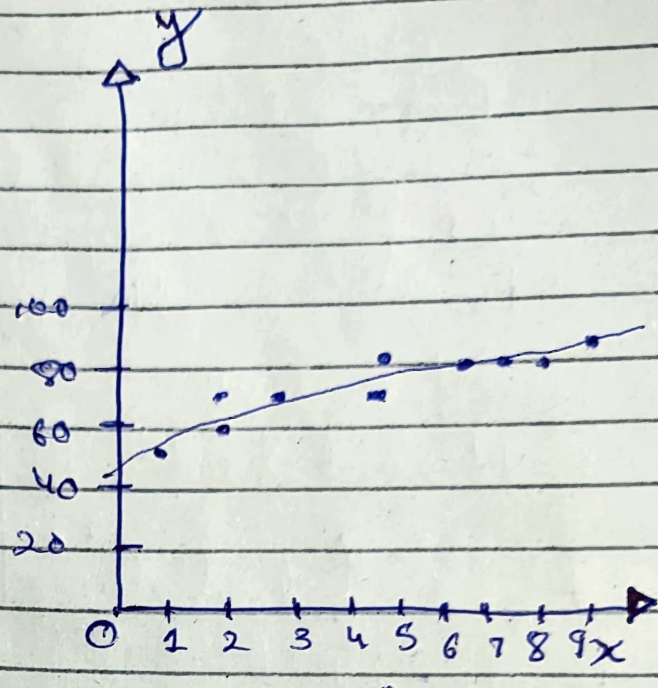
Optimisation

↓

model

Simple Linear Regression

Hours x	Marks y
2	69
9	98
5	82
5	77
3	71
7	84
1	55
8	94
6	84
2	64



Error $= y - \hat{y}$

Error = Actual value − predicted Value

$$\boxed{e_s = \sum_{i=1}^{n} (y - \hat{y})^2}$$ ✓ $\hat{y}_i = b_0 + b_1 x_i$

OR ⟶ Cost function

$$\boxed{e_a = \sum_{i=1}^{n} |y_i - \hat{y}_i|}$$

$$\cancel{e_s = \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_i}$$

$$e_s \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_i)]^2$$

$$\boxed{\text{Min} \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_i)]^2}$$

$$\underset{b_0, b_1}{\arg\min} \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_i)]^2$$

least squares optimization sum of squares of errors.

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$\bar{x}$ = avg of x

$(2 - \text{avg}x)(69 - \text{avg}y)$
$(9 - \text{av})(98 - \text{avg}y)$

$$b_0 = \bar{y} - b_1 \bar{x}$$

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
2	69	-2.8	-8.8	24.64	7.84
9	98	4.2	20.2		
5	82	0.2	4.2		
5		0.2	-0.8		
3	77 / 71	1.8	-6.8		
7	84	2.2	6.2		
1	55	-3.8	-22.8		
Σ 48	778			320.6	67.6
$\bar{x} = 4.8$	$\bar{y} = 77.8$				
8	94	3.2	16.2		
6	84	1.2	6.2		
2	64	-2.8	-13.8		

$$b_1 = \frac{320.6}{67.6} = 4.74$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_0 = 77.8 - (4.74)(4.8)$$

$$b_0 = 55.048$$

$$\boxed{\hat{y}_i = 55.048 + 4.74\, x_i}$$

$x = 4$

$$\boxed{\hat{y}_i = 74.008}$$

$x = 3$

$$\hat{y}_i = 69.268$$

y

...g Bayes Theorem

AI

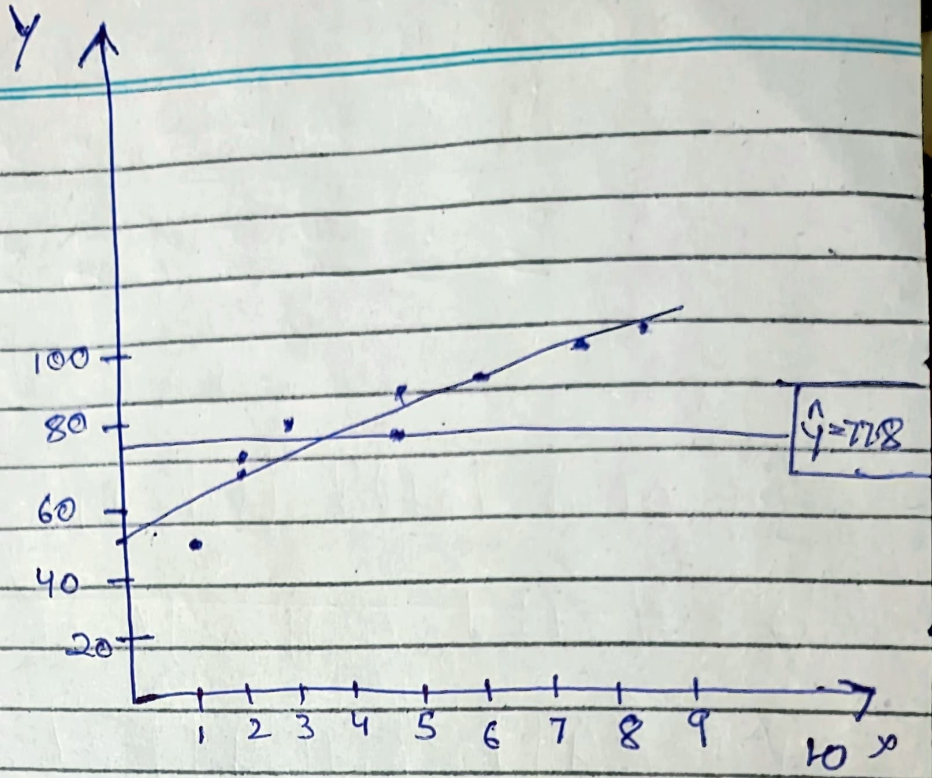
Simple linear Regression: Analyze specific relationships between 2 or more variables.

Dependent variable $\Rightarrow y = a_1x_1 + a_2x_2 + \cdots + a_n x_n$

$x = [x_1, x_2, \ldots, x_n]$

[x is independent]

x	y
2	69
9	98
5	82
5	77
3	71
7	84
1	55
8	94
6	84
2	64



$$y = 55.04 + 4.74x$$

x	y
0	55.04
2	64.5
4	73.96
6	83.48
8	92.96
10	102.4

$x = 5.5$

$y = 55.04 + 4.74(5.5)$

$\boxed{y = 81.99}$

Coefficient of Determination

$$r^2 = \frac{SSR}{SST}$$

SSR = Sum of Squares due to regression

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

SST = Sum of Squares of total deviation

$$= \sum(y_i - \bar{y})^2$$

SSE = Sum of Squares of Errors

$$= \sum(y_i - \hat{y}_i)^2$$

$$\boxed{SST = SSR + SSE}$$

x_i	y_i	$\hat{y}_i = b_0 + b_1 x$	Error $y_i - \hat{y}_i$	squared error $(y_i - \hat{y}_i)^2$	Derivation $y_i - \bar{y}$	squared Deviation $(y_i - \bar{y})^2$
2	69					
9	98					
5	82					
5	77					
3	71					
7	84					
1	55					
8	94					
6	84					
2	64					
				SSE = 79.13		SST 1599.6

$SST = SSR + SSE$

$\boxed{SSR = 1599.6 - 79.13 = 1520.47}$

$SSR = SST - SSE$

$\boxed{r^2 = \frac{SSR}{SST}}$

Coefficient of determination r^2 measures the percent of variability in y that can be explained by $\hat{y}$.

Correlation coefficient

$r_{xy} = (\text{sign of } b_1)\sqrt{r^2}$

$r^2 = \frac{1520.47}{1599.6} = 0.9505$

$\boxed{r^2 = 0.9505}$

$r_{xy} = +\sqrt{0.9505}$

$r_{xy} = 0.9749$

AI formulas

$\bar{x} = \frac{\sum xi}{n(xi)}$ $\bar{y} = \frac{\sum yi}{n(yi)}$

$\rightarrow \hat{y} = b_0 + b_1 xi$ — $b_0 = \bar{y} - b_1 \bar{x}i$

$SSE = \sum (y_1 - \hat{yi})^2$

$SSR = \sum (\hat{yi} - \bar{y})^2$ $r^2 = \frac{SSR}{SSE}$

$SST = \sum (yi - \bar{y})^2$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_1 - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$r_{xy} = (\text{sign of } b_1)(\sqrt{r^2})$

Machine Learning pipeline

1// Problem Definition

2// Data Ingestion

3// Data preparation.

↳ Filter, clean, missing value, duplicate

4// Data Segregation → training set, test

5// Model training → Regression, Classification, Clustering

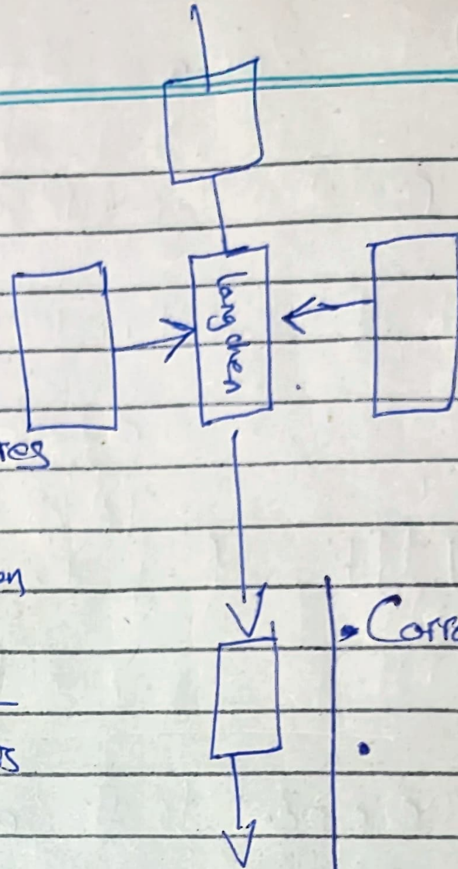
6// Model Evaluation

7// Model Deployment

8// Performance Monitoring

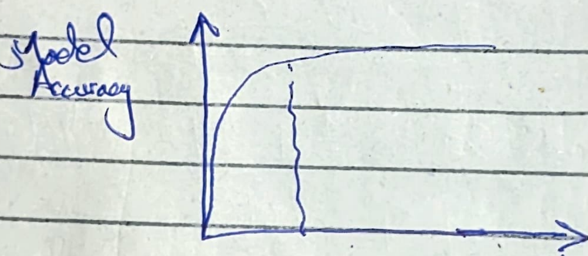
Feature Selection

- Variable selection
- Attribute
- Relevant features
- Correct
- Reasons to drop features
  - ↳ Irrelevancy
  - ↳ ~~Correlated~~ High Correlation
  - ↳ Redundant features



- Correlation heatmap
-

Exploratory data Analysis



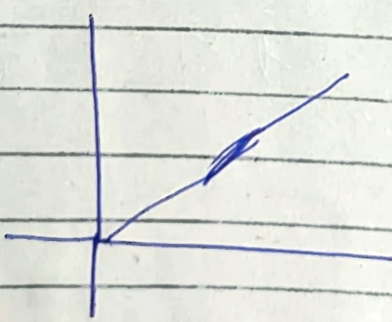
$$\rho_{x,y} = \frac{Cov[x,y]}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$= \frac{Cov[x,y]}{\sqrt{VAR[x]\,VAR[y]}}$$

$$Cov[x,y] = E[xy] - E[x]E[y]$$

$$VAR[x] = E[x^2] - (E[x])^2$$

$\bar{x}\bar{y}$



	A	B	C
A	2	5	4
B	1	5	6
C	4	9	7

Mean

Stand Deviation

Covariance

Correlation

effect

A1. Data preprocessing Techniques

1. Binarization
2. Data Binning
3. Mean Removal
4. Scaling
   (a) Min-max scaling or normalization
   (b) z-score scaling or standardization

Binarization.

1. Binary classification
2. feature encoding
3. Text data Binarization
4. Image Thresholding
5. Anomaly detection

1. Binary Classification: 0 or 1 (Positive or Negative)

2. Feature Encoding:

Categorical or nominal features

Car	Red	Green	Blue
1	1	0	0
2	0	1	0
3	1	0	0
4	0	0	1
5	0	1	0

One-hot encoding

00 Red
01 Green
10 Blue
11

Binary Encoding

AI

Hamming distance:

0 1 0
↕ ↕ ↕
0 1 1

$d(010, 011) = 1$

XOR

X	Y	X ⊕ Y
0	0	0
0	1	1
1	0	1
1	1	0

```
 0110  0101
 1001  0110
 ----------
 1111  0011
```

$d = 6$

Text Data Binarization:

vocabulary: {apple, banana, cherry, date}

Document 1: Amjad likes Apples and Cherries

Document 2: Bananas are rich in magnesium.

	apple	Banana	Cherry	date	
Document 1:	1	0	1	0	{1,0,1,0}
Document 2:	0	1	0	0	{0,1,0,0}

Image Thresholding:

$4 \geq \boxed{1}$
$4 < \boxed{0}$

7 8 9 6
3 0 6 3
2 5 1 2 ⟹
4 3 7 9

1	1	1	1
0	0	1	0
0	1	0	0
1	0	1	1

- feature extraction
- Boundary detection
- Compression

Anomaly Detection

Unusual patterns outliers

- Define Threshold
- Binarization
- Detect anamoly

~~Threshold = 1000 bytes~~
~~Packet size~~

Threshold = 1000 bytes

	Packet Size	Binary Value	Status
(1)	1200	0	Normal
(2)	1200	1	anomaly
(3)	1400	0	Normal
(4)	2000	1	anomaly

AI

Data Binning

$$b_0 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}, \quad b = \bar{y} - b_1 \bar{x}$$

$$\hat{y}_i = b_0 + b_1 x_i \qquad SST = \sum (y_i - \bar{y})^2$$

~~SS~~ Data Binning:

Data discretization / Bucketing:

Example

Ages: {25, 30, 32, 45, 55, 60, 62, 70, 75, 80}

Bin 1: [21, 30] Young adults → {25, 30}

Bin 2: [31, 40] Adults → {32}

Bin 3: [41 - 50] Middle aged → {45}

Bin 4: [51, 60] Seniors → {55, 60}

Bin 5: [61, 70] Elders → {62, 70}

Bin 6: [71, 80] very Elderly → {75, 80}

Scores = {150, ~~1~~70, 112, 84, 32, 0, 16, 53, 41, 22}

Bin Width = 50

Bin 1: [0, 50] low → {32, 0, 16, 41, 22}

Bin 2: [50, 100] Medium → {70, 84, 53}

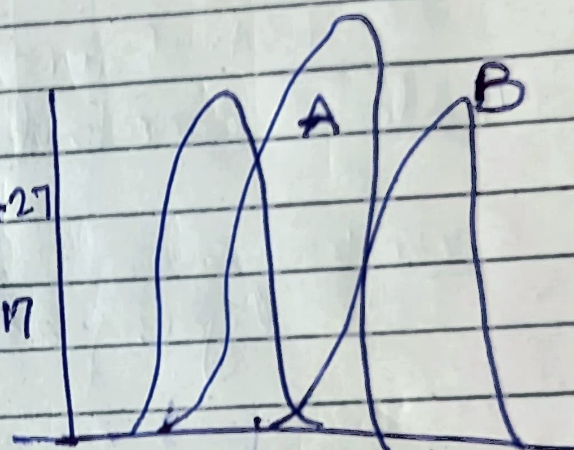
Bin 3: [100, 150] High → {150, 112}

Mean Removals

	A	B	C	
15 < 3/4 20, 20	80	~~60~~ 77	55	427
	60	53	82	
15	75	92	65	17

Sum = 215 222 202

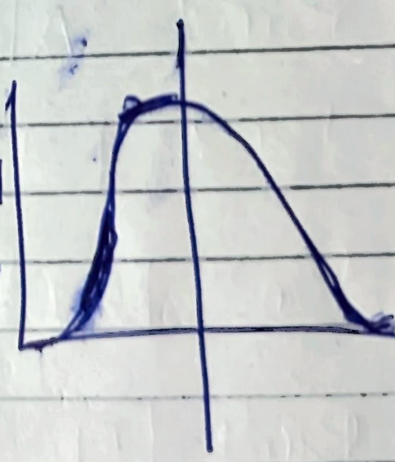
Mean = 71.67 74 67.33



	A	B	C	
20	8.33	3	-12.33	427
	-11.67	-21	14.67	
15	3.33	18	-2.33	17

Sum = -0.01

Mean = -0.003 0 0.003



AI Data Preprocessing:

(a) Min-max Scaling / Normalization
(b) z-score / standardization

Min-Max Scaling

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

	A	B	C
	80	77	55
	60	53	82
	75	92	65
X_{min}	60	53	55
X_{max}	80	92	82
$X_{max} - X_{min}$	20	39	27

A

$\frac{80-60}{20} = 1$

0

$\frac{75-60}{20} = 0.75$

B

$\frac{77-53}{\cancel{20}}$

A	B	C
1	0.61	0
0	0	1
0.75	1	0.37

(b) Z-score scaling:

$Z = \frac{x-\mu}{S}$

S = Standard Deviation

$S^2 = \frac{1}{n-1} \sum (x-\bar{x})^2$ Variance

α	β	γ	η
12	5	0.7	70
7	-5	0.3	65
8	0	0.5	30
15	6	0.9	90
10.5	1.5	0.6	63.75 = Mean

VAR = 13.67 | 25.667 | 0.068 | 622.9

STD = 3.697 | 5.06 | 0.258 | 24.96

$Z = \frac{12-10.5}{\cancel{8}3.697}$

Scaling

$\bar{\alpha}$	$\bar{\beta}$	$\bar{\gamma}$	$\bar{m}$

Relative Distance ?

Gradient Ascent

AI Gradient Descent

Gradient Ascent.

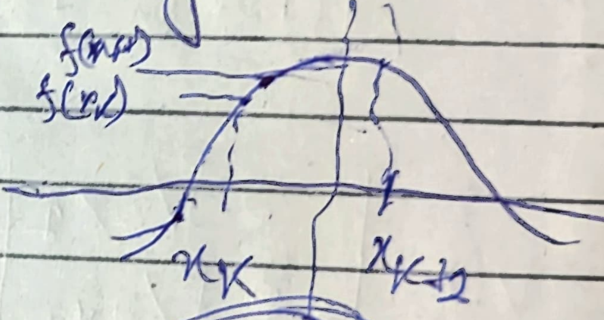
1. Choose a starting point, initial vector
2. Perform Iteration to get to next point

$$\vec{x}_{K+1} = \bar{x}_K + \lambda_K^* \nabla f \big|_{x_K}$$

↳ step size

↳ learning rate

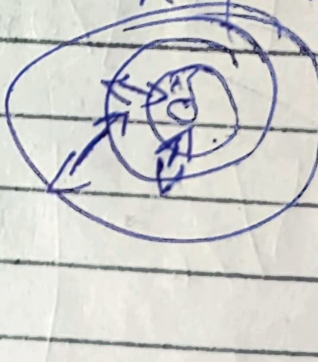
3. $$\left| f(x_{K+1}) - f(x_K) \right| < \epsilon$$



Maximum

Gradient Descent

$$\boxed{x_{K+1} = x_K - \lambda_K^* \nabla f \big|_{x_K}}$$



Contours Plot

$$\text{Min } f(x)$$
$$\text{Max } -f(x)$$

Minimize $Z = (x_1 - \sqrt{5})^2 + (x_2 - \pi)^2 + 10$

$X_0 = [6.597, 5.891]^T$ (x_1, x_2) $\quad \epsilon = 0.05$

Maximize $Y = -(x_1 - \sqrt{5})^2 - (x_2 - \pi)^2 - 10$

$f(x_0) = -(6.597 - \sqrt{5})^2 - (5.891 - \pi)^2 - 10$

$f(x_0) = -36.577$

Iteration 1

$x_1 = x_0 + \lambda_1 \nabla f|_{x_0}$

$$\nabla f = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \end{bmatrix} = \begin{bmatrix} -2(x_1 - \sqrt{5}) \\ -2(x_2 - \pi) \end{bmatrix}$$

$$\nabla f/_{x_0} = \begin{bmatrix} -2(6.597 - \sqrt{5}) \\ -2(5.891 - \pi) \end{bmatrix} = \begin{bmatrix} -8.722 \\ -5.499 \end{bmatrix}$$

$$x_1 = \begin{bmatrix} 6.597 \\ 5.891 \end{bmatrix} + \lambda_1 \begin{bmatrix} -8.722 \\ -5.499 \end{bmatrix}$$

$$\vec{x_1} = \begin{bmatrix} 6.597 - 8.722\lambda_1 \\ 5.891 - 5.499\lambda_1 \end{bmatrix} \rightarrow (1)$$

$$f(\vec{x_1}) = -(6.597 - 8.722\lambda_1 - \sqrt{5})^2 - (5.891 - 5.499\lambda_1 - \pi^2) - 10$$

$$f(\bar{x}_1) = -106.312\lambda_1^2 + 106.307\lambda_1 - 36.575$$

Take first derivative

$$f'(\bar{x}_1) = -2(106.312)\lambda_1 + 106.307 = 0$$

$\lambda_1 = \frac{1}{2} = 0.5 \rightarrow$ Stationary point or Critical point

$$f''(\bar{x}_1) = -2(106.312) < 0 \leftarrow \boxed{\text{Maximum}}$$

Using $\lambda_1 = 0.5$ in ①

$$\bar{x}_1 = \begin{bmatrix} 6.597 - 8.722(0.5) \\ 5.891 - 5.499(0.5) \end{bmatrix}$$

$$\bar{x}_1 = \begin{bmatrix} 2.236 \\ 3.142 \end{bmatrix} \qquad f(x_1) = ?$$

$$f(\bar{x}_1) = -10 \rightarrow 9.9995 = -10$$

$$|f(\bar{x}_1) - f(\bar{x}_0)| < \epsilon = 0.05$$

$$-10 + 36.577 > 0.05$$

we need another iteration

Iteration 2:

$$\bar{x}_2 = \bar{x}_1 + \lambda_2 \nabla f|_{\bar{x}_1}$$

AI: Continued 2nd Iteration:

Second Iteration

$$\vec{x}_2 = \vec{x}_1 + \lambda_2 \nabla f|_{x_1}$$

$$\vec{x}_1 = \begin{bmatrix} 2.236 \\ 3.142 \end{bmatrix}$$

$$\nabla f|_{x_1} = \begin{bmatrix} -2(2.236-\sqrt{5}) \\ -2(3.142-\pi) \end{bmatrix} = \begin{bmatrix} 0.0001 \\ -0.0008 \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\vec{x}_2 = \vec{x}_1 + \lambda_2 \nabla f|_{x}$$

$$= \begin{bmatrix} 2.236 \\ 3.142 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\vec{x}_2 = \begin{bmatrix} 2.236 \\ 3.142 \end{bmatrix} = \vec{x}_1$$

$$f(\vec{x}_2) = f(\vec{x}_1) = -10$$

$$|f(\vec{x}_2) - f(\vec{x}_1)| = -10+10 = 0 < \epsilon = 0.05$$

$$\vec{x}^* = \begin{bmatrix} 2.236 \\ 3.142 \end{bmatrix}$$

$Z^* = -10$

The given function $f(x)$ is minimum at

$$x^* = [2.236 \;\; 3.142]^T$$

$Z^* = -10$

The given function $f(x)$ is minimum at

$$X^* = [2.236 \quad 3.142]^T$$

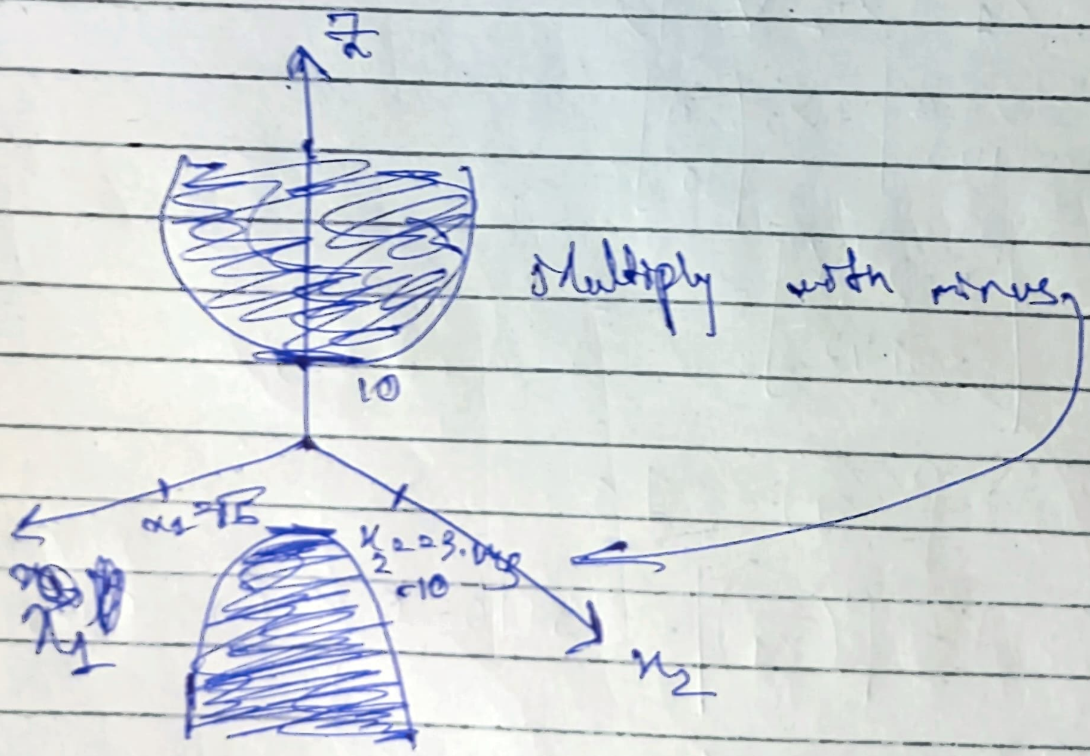
$$Z^* = 10 \quad \text{(Minimum)}$$

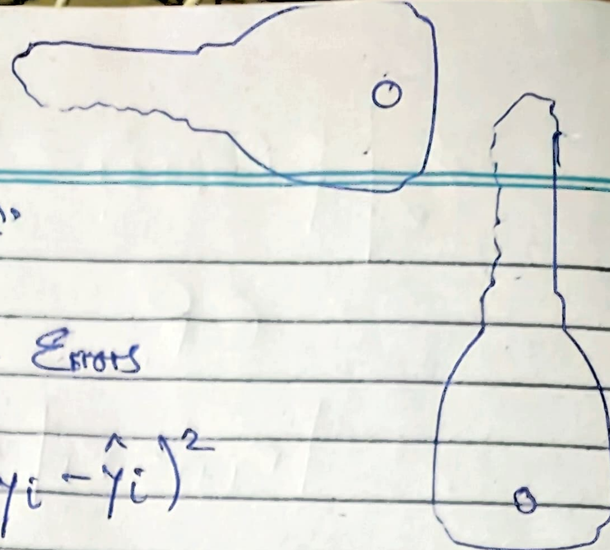
$$\lambda = 0.5$$

$$X_2 = \begin{bmatrix} 2.236 \\ 3.142 \end{bmatrix} + 0.5 \begin{bmatrix} 0.0001 \\ -0.0008 \end{bmatrix}$$

$$\vec{X}_2 = \vec{X}_1$$

$$Z = (x_1 - \sqrt{5})^2 + (x_2 - \pi)^2 + 10$$





AI Cost function Minimization.

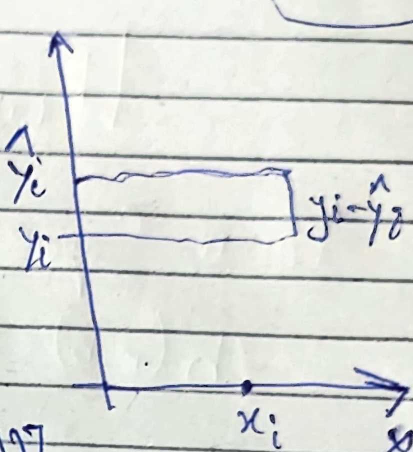
1- Sum of Squares of Errors

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

$\hat{y}_i = \hat{m} x_i + \hat{b}$

$$Z = \sum_{i=1}^{n} \left[y_i - (\hat{m} x_i + \hat{b})\right]^2$$

$$f(m,b) = \sum_{i=1}^{n} \left[y_i - (\hat{m} x_i + \hat{b})\right]^2$$



$$\frac{\partial f}{\partial m} = -2 \sum_{i=1}^{n} x_i \left[y_i - (\hat{m} x_i + \hat{b})\right]$$

$$\frac{\partial f}{\partial b} = -2 \sum_{i=1}^{n} \left[y_i - (\hat{m} x_i + b)\right]$$

x_i	y_i
2	58
3	60
5	70
7	80

$$\nabla f = \begin{bmatrix} \partial f / \partial m \\ \partial f / \partial b \end{bmatrix}$$

1- Starting Point

$$w_0 = [m_0 \quad b_0]^T$$

2- Iteration: $w_{i+1} = w_i - \lambda_i \nabla f|_{w_i}$

i.e $m_{i+1} = m_i + \alpha \frac{\partial f}{\partial m}\Big|_{m_i}$

$b_{i+1} = b_i + \alpha \frac{\partial f}{\partial b}\Big|_{b_i}$

$$3/\checkmark \; |f(w_{i+1}) - f(w_i)| < \epsilon$$

OR

$$|m_{i+1} - m_i| < \epsilon$$

$$|b_{i+1} - b_i| < \epsilon$$

Mean Square Error (MSE)

$$h(m,b) = 1/n \sum_{i=1}^{n} [y_i - (\hat{m} x_i + \hat{b})]^2$$

$$\frac{\partial h}{\partial m} = -2/n \sum_{i=1}^{n} x_i [y_i - (\hat{m} x_i + \hat{b})]$$

$$\frac{\partial h}{\partial b} = -2/n \sum_{i=1}^{n} [y_i - (\hat{m} x_i + \hat{b})]$$

$$\nabla h$$

3- Mean Absolute Error:

$$g = 1/n \sum_{i=1}^{n} |y_i - (\hat{m} x_i + \hat{b})|$$