Simple Linear Regression

$f(i/p) \longrightarrow o/p$ Mapping

Eg: $f(\text{House Area, No. of bedroom}) \longrightarrow$ House Price

- It is a curve fitting problem.

* Partial derivative →

- $f(x,y) = 3x^2y - 2 + y^3$

$$\frac{df}{dx} = 6xy$$

$$\frac{df}{dy} = 3x^2 + 3y^2$$

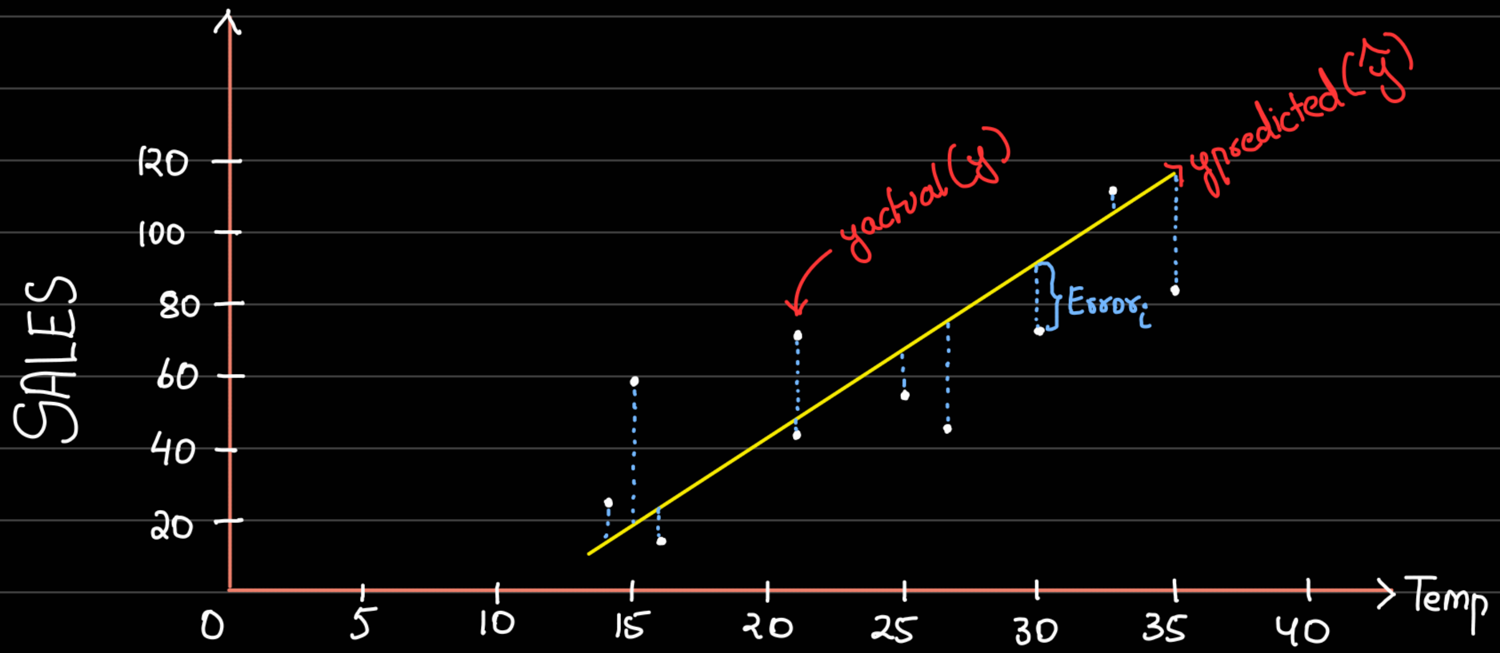
- $f(x,y) = 2x + 3y - 4$

$$\frac{df}{dx} = 2$$

$$\frac{df}{dy} = 3$$

* Ice cream sales data →

Sales	Temp
42	21
18	16
25	14
74	30
112	32
71	21
58	25
46	27
82	35
59	15



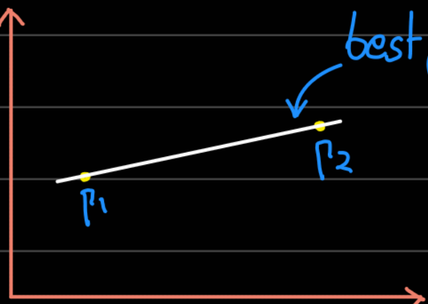
$$\text{Error}_i\ (E_i) = \text{True Value} - \text{Predicted value}$$
$$= y_i - \hat{y}_i$$

Best fit line $\hat{y}_i = a + b X_i$

intercept (a) slope (b)

Best line = For a line for which we have $e_1, e_2, e_3 \ldots e_n$ minimum $\rightarrow$

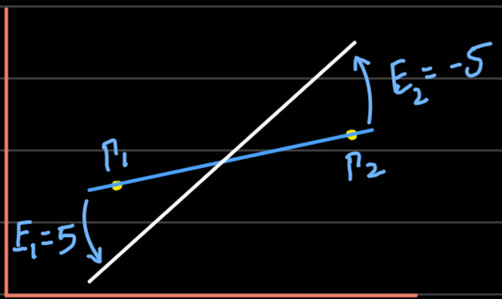
$$\sum_{i=1}^{n} E_i \longrightarrow \text{minimum}$$



$$\boxed{\sum_{i=1}^{2} \varepsilon_i = 0}$$

$\downarrow$ If our line was

I Scenario - 1 : $\sum_{i=1}^{n} E_i \rightarrow \min$



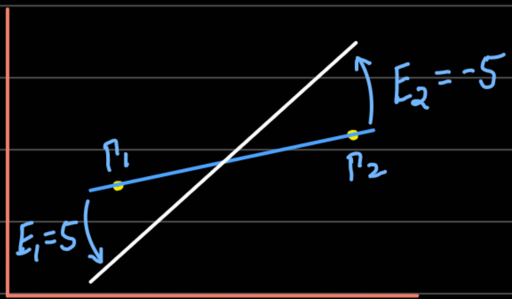
$y_1 = P_1 = 10$, $\hat{y}_1 = 5$
$E_1 = 10 - 5 = 5$

$y_2 = P_2 = 20$, $\hat{y}_2 = 25$
$E_2 = 20 - 25 = -5$

Now $\sum_{i=1}^{2} E_i = 5 - 5 = 0$ (This fitted line also gives error as minimum)

So, the metric to minimize $\sum E_i$ is not successful.

II Scenario - 2 : $\sum_{i=1}^{n} |E_i| \longrightarrow \min$



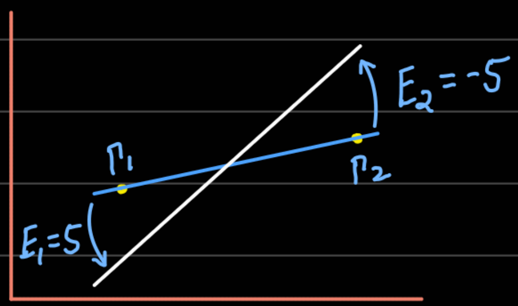
$y_1 = P_1 = 10$, $\hat{y}_1 = 5$
$E_1 = 10 - 5 = 5$

$y_2 = P_2 = 20$, $\hat{y}_2 = 25$
$E_2 = 20 - 25 = -5$

$$\text{Now, } \sum_{i=1}^{2} |E_i| = |E_1| + |E_2| = 5+5 = \underline{\underline{10}}$$

This metric is giving best line, but the problem is, it is difficult to differentiate. A ML model needs a metric which is easy differentiable to optimize (smooth function)

III Scenario - 3 : $\sum_{i=1}^{n} E_i^2 \rightarrow \min$



$y_1 = P_1 = 10$, $\hat{y}_1 = 5$
$E_1 = 10 - 5 = 5$

$y_2 = P_2 = 20$, $\hat{y}_2 = 25$
$E_2 = 20 - 25 = -5$

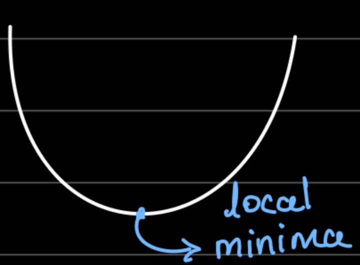
$$\text{Now, } \sum_{i=1}^{2} E_i^2 = E_1^2 + E_2^2 = 25+25 = 50$$

This metric is giving best line, as well as it is differentiable, so it works as required.

* Optimal value of a, b in $y = a + bx$

We know, $E_i = (y_i - \hat{y}_i)^2$

$$\underbrace{\sum_{i=1}^{n} E_i}_{E} = \sum_{i=1}^{n} (y_i - (a+bx))^2 \rightarrow \text{To minimize}$$



This is equation of parabola, so after partial differentiation no need to check maxima, minima saddle point as it only have single optimal point i.e. local minima.

(i) $\frac{\partial E}{\partial a} = 0$

$$\Rightarrow \frac{\partial}{\partial a}\left(\sum_{i=1}^{n}\left(y_i - (a+bx_i)\right)^2\right)$$

$$\Rightarrow 2\sum_{i=1}^{n}\left(y_i - (a+bx_i)\right)(-1) = 0$$

$$\Rightarrow \sum_{i=1}^{n}\left(y_i - (a+bx_i)\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - b\sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i - na - b\sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i$$

Dividing equation by n

$$\Rightarrow \frac{\sum_{i=1}^{n} y_i}{n} = \frac{na}{n} + \frac{b\sum_{i=1}^{n} x_i}{n}$$

$$\Rightarrow \boxed{\bar{Y} = a + b\bar{X}} \qquad \text{mean}(y) = a + b \cdot \text{mean}(x)$$

eq ①

(ii) $\frac{\partial E}{\partial b} = 0$

$$\Rightarrow \frac{\partial}{\partial b}\left(\sum_{i=1}^{n}\left(y_i - (a+bx_i)\right)^2\right)$$

$$\Rightarrow 2\sum_{i=1}^{n}\left(y_i - (a+bx_i)\right) * (-x_i) = 0$$

$$\Rightarrow -\sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n}(ax_i + bx_i^2) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n}(ax_i + bx_i^2)$$

Dividing equ by n

$$\Rightarrow \frac{\sum_{i=1}^{n} x_i y_i}{n} = \frac{\sum_{i=1}^{n}(ax_i + bx_i^2)}{n}$$

$$\Rightarrow \boxed{\overline{XY} = a\,\overline{X} + b\,\overline{X^2}}$$

eq 2

mean $(x \cdot y) = a \cdot$ mean $(x) + b \cdot$ mean (x^2)

We got 2 eqns $\rightarrow$ $\left.\begin{aligned} \overline{Y} &= a + b\overline{X} \\ \overline{XY} &= a\overline{X} + b\overline{X^2} \end{aligned}\right\}$ Analytical (Closed-form) Solution.

We can write these eqns in form of $AX = B$ (System of eqns)

$$\underbrace{\begin{bmatrix} 1 & \overline{X} \\ \overline{X} & \overline{X^2} \end{bmatrix}}_{A} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \overline{Y} \\ \overline{XY} \end{bmatrix}$$

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & \overline{X} \\ \overline{X} & \overline{X^2} \end{bmatrix}^{-1} \begin{bmatrix} \overline{Y} \\ \overline{XY} \end{bmatrix}$$

$$\boxed{\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{|A|} \begin{bmatrix} \overline{X^2} & -\overline{X} \\ \overline{X} & 1 \end{bmatrix} \begin{bmatrix} \overline{Y} \\ \overline{XY} \end{bmatrix}}$$

* Deriviative with respect to vector

I linear Function

$$f(\omega) = \alpha^T \omega + \beta$$

$$= [\quad]^T_{1\times n} [\quad]_{n\times 1} + \beta = [\quad]_{1\times 1}$$

Scalar value

$$= [\alpha_1 \ \alpha_2 \ \alpha_3 \ldots \alpha_n] \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{bmatrix} + \beta$$

$$f(\omega) = \alpha_1 \omega_1 + \alpha_2 \omega_2 + \alpha_3 \omega_3 \cdots + \alpha_n \omega_n + \beta$$

$$\nabla f(\omega) = \begin{bmatrix} \partial f / \partial \omega_1 \\ \partial f / \partial \omega_2 \\ \vdots \\ \partial f / \partial \omega_n \end{bmatrix} \qquad \begin{matrix} \partial f / \omega_1 = \alpha_1 \\ \partial f / \omega_2 = \alpha_2 \\ \vdots \\ \partial f / \omega_n = \alpha_n \end{matrix}$$

$$= \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \underline{\underline{\alpha}}$$

$$\boxed{\text{So if } f(\omega) = \alpha^T \omega + \beta, \text{ then } \nabla f(\omega) = (\alpha^T)^T = \underline{\underline{\alpha}}}$$

II Quadratic function →

$$f(\omega) = \omega^T A \omega$$

$$\boxed{\nabla f(\omega) = (A^T + A)\,\omega}$$

Calculation :-

$$f(\omega) = \underset{1 \times n}{\begin{bmatrix} \omega_1 & \omega_2 & \omega_2 & \cdots & \omega_n \end{bmatrix}} \underset{n \times n}{\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ X_{n1} & \cdots & \cdots & X_{nn} \end{bmatrix}} \underset{n \times 1}{\begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{bmatrix}}$$

$$f(w) = \begin{bmatrix} t_1 & t_2 & t_3 & \cdots & t_n \end{bmatrix}_{n\times 1} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}_{n\times 1} \quad \text{where } t_i = \sum_{j=1}^{n} w_j x_{ji}$$

$$f(w) = \left[t_1 w_1 + t_2 w_2 + \cdots + t_n w_n \right]$$

$$\nabla f(w) = \begin{bmatrix} \partial f/\partial w_1 \\ \partial f/\partial w_2 \\ \partial f/\partial w_3 \\ \vdots \\ \partial f/\partial w_n \end{bmatrix}$$

$$\frac{\partial f}{\partial w_1} = \frac{\partial (t_1 w_1 + t_2 w_2 + \cdots + t_n w_n)}{\partial w_1}$$

$$= \frac{\partial t_1 w_1}{\partial w_1} + \frac{\partial t_2 w_2}{\partial w_1} + \cdots + \frac{\partial t_n w_n}{\partial w_1}$$

$$= \frac{\partial (w_1 x_{11} + w_2 x_{21} + \cdots + w_n x_{n1}) w_1}{\partial w_1} +$$

$$\frac{\partial (w_1 x_{12} + w_2 x_{22} + \cdots + w_n x_{n2}) w_2}{\partial w_1} +$$

$$\vdots$$

$$\frac{\partial (w_1 x_{1n} + w_2 x_{2n} + \cdots + w_n x_{nn}) w_n}{\partial w_1}$$

$$= \left(2w_1 x_{11} + w_2 x_{21} + \cdots + w_n x_{n1}\right) + \left(x_{12} w_2\right) + \left(x_{13} w_3\right) + \cdots + \left(x_{1n} w_n\right)$$

$$= (x_{11} + x_{11}) w_1 + (x_{12} + x_{21}) w_2 + (x_{13} + x_{31}) w_3 + \cdots + (x_{1n} + x_{n1}) w_n$$

$$= \sum_{k=1}^{n} \underbrace{(x_{1k} + x_{k1})}_{\downarrow} w_k$$

Summation of First row of A and its transpose

$$= \left(\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \end{bmatrix} + \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \end{bmatrix}\right) \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

Similarly,

$$\frac{\partial f}{\partial w_2} = \sum_{k=1}^{n} (x_{2k} + x_{k2}) w_k$$

$$\frac{\partial f}{\partial w_3} = \sum_{k=1}^{n} (x_{3k} + x_{k3}) w_k$$

$$\vdots$$

$$\frac{\partial f}{\partial w_n} = \sum_{k=1}^{n} (x_{nk} + x_{kn}) w_k$$

$$\nabla f(w) = \begin{bmatrix} \partial f / \partial w_1 \\ \partial f / \partial w_2 \\ \partial f / \partial w_3 \\ \vdots \\ \partial f / \partial w_n \end{bmatrix}$$

$$= \left(\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ X_{n1} & \cdots & \cdots & X_{nn} \end{bmatrix} + \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ X_{1n} & \cdots & \cdots & X_{nn} \end{bmatrix} \right)_{n \times n} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}_{n \times 1}$$

$$= (A + A^T) w \quad \text{or} \quad (A^T + A) w$$

Given the scalar function:

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$$

Where:

- $\mathbf{x}$ is a vector,
- $\mathbf{Q}$ is a matrix,
- $\mathbf{b}$ is a vector,
- c is a scalar constant.

Find the derivative of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

$$f = \frac{1}{2} x^T Q x + b^T x + c$$

$$\frac{df}{dx} = \frac{1}{2} (Q^T + Q) x + (b^T)^T + 0$$

$$= \frac{x}{2} (Q^T + Q) + b$$

Given the scalar function

$$f(\mathbf{x}) = \mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{A}\mathbf{x}$$

Where:

- $\mathbf{Y}$ is a vector,
- $\mathbf{A}$ is a matrix.

Find the derivative of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

$$f = Y^T Y - Y^T A x$$

$$\frac{df}{dx} = 0 - (Y^T A)^T$$

$$\frac{df}{dx} = -A^T Y$$

$$f(\mathbf{x}) = -\mathbf{x}^T \mathbf{A}^T \mathbf{Y}$$

Where:

- $\mathbf{x}$ is a vector,
- $\mathbf{A}$ is a matrix,
- $\mathbf{Y}$ is a vector.

Find the derivative of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

$f = -x^T A^T Y$

$\frac{df}{dx} = -A^T Y$

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x}$$

Where:

- $\mathbf{x}$ is a vector,
- $\mathbf{A}$ is a matrix.

Find the derivative of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

$f = x^T A^T A x$

$\frac{df}{dx} = A^T A + (A^T A)^T$

$= (A^T A + A^T A) x$

$\frac{df}{dx} = 2 A^T A x$

$$f(\mathbf{x}) = (\mathbf{Y} - \mathbf{A}\mathbf{x})^T (\mathbf{Y} - \mathbf{A}\mathbf{x})$$

Where:

- $\mathbf{Y}$ is a vector,
- $\mathbf{A}$ is a matrix,
- $\mathbf{x}$ is a vector.

Find the derivative of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

$f = (Y - Ax)^T \cdot (Y - Ax)$

$$f = (Y^T - x^T A^T)(Y - Ax)$$
$$= Y^T Y - Y^T A x - x^T A^T y + x^T A^T A x$$

$$\frac{df}{dx} = 0 - (Y^T A)^T - (A^T Y) + (A^T A + (A^T A)^T) x$$
$$= - A^T Y - A^T Y + (A^T A + A^T A) x$$
$$= -2A^T Y + 2A^T A x$$
$$= 2(-A^T Y + A^T A x)$$

Question: In regression analysis, what is the nature of the output variable?

- A) Discrete
- B) Continuous and confined within a specific range
- **C) Continuous**
- D) Can be either discrete or continuous

Continuous.

Consider a regression line $y = ax + b$, fitted from a set of numbers, where a is the slope and b is the intercept.
If we know the value of the slope a, then by using which option can we always find the value of the intercept b?

A. Put the value (0,0) in the regression line.
B. Put any value from the points used to fit the regression line and compute the value of b.
C. Put the mean values of x and y in the equation along with the value a to get b.
D. None of the above.

We know : $\bar{Y} = a\bar{X} + b$

$b = \bar{Y} - a\bar{X}$

(C) Ans

Question: Consider the linear regression model $y = a + bx$ where the mean value of the independent variable x is 3.00 and the mean value of the dependent variable y is 4.00, with $a = 2.00$. Given these values, what is the correct value for the slope parameter b of the model?

$\overline{X} = 3.00$ $\quad \overline{Y} = 4.00$ $\quad a = 2.00$

$$\overline{Y} = a + b\overline{X}$$
$$4 = 2 + 3b$$
$$\boxed{b = \frac{2}{3}}$$ Ans.

Consider a dataset with values $X = [1, 2, 3, 4]^T$ and $Y = [3, 4, 8, 11]^T$. The prediction model is defined as $\hat{y} = 2x + 1$. Calculate the mean squared error (MSE) for this model and choose the correct answer from the options below:

A. 0.75

B. 1.5

C. 2.5

D. 3.0

X	Y
1	3
2	4
3	8
4	11

$$MSE = \frac{1}{n}(y - \hat{y})^2 = \frac{1}{n}(y - 2x - 1)^2 = \frac{\cancel{6}\,\cancel{3}\,1.5}{\cancel{4}\,\cancel{2}} = \underline{\underline{1.5}} \text{ Ans.}$$

x	y	$2x$	$y-2x-1$	ε_i^2
1	3	2	0	0
2	4	4	-1	1
3	8	6	1	1
4	11	8	2	4
				6

Consider the dataset from a study investigating the correlation between the number of hours spent driving and the risk of developing acute back pain. The dataset includes pairs of driving hours (x) and corresponding risk scores (y) on a scale of 0-100.

Fit a best-fit line to this data and using the best-fit line, estimate the risk score for 20 hours of driving. Which of the following values is closest to the predicted risk score?

A. 79.42
B. 108.38
C. 104.38
D. 82.38

Number of Hours Spent Driving (x)	Risk Score (y)
10	95
9	80
2	10
15	50
10	45
16	98
11	38
16	93

X	Y	XY	X^2
10	95	950	100
9	80	720	81
2	10	20	4
15	50	750	225
10	45	450	100
16	98	1568	256
11	38	418	121
16	93	1488	256
11.125	63.625	795.5	142.875

$\overline{Y} = a\overline{X} + b$ $\qquad$ $63.625 = 11.125a + b$

$\overline{XY} = a\overline{X^2} + b\overline{X}$ $\qquad$ $795.5 = 142.875a + 11.125b$

$a = 4.588$ $\qquad$ $b = 12.584$

$y = 4.588x + 12.584 \Rightarrow 4.588(20) + 12.584 = \underline{\underline{104.344}}$ Ans.

Consider the following dataset where the mass y (grams) of a chemical is related to the time x (seconds), for which the chemical reaction has been taking place according to the table:

Time, x (seconds)	Mass, y (grams)
5	40
7	120
12	180
16	210
20	240

Q1: What is the mass of the chemical after ten seconds has passed?

A. 61.40
B. 133.58
C. 121.71
D. 98.30

Q2: By how much does the chemical increase in weight in five seconds?

A. 61.40
B. 133.58
C. 121.71
D. 98.30

X	Y	XY	X^2
5	40	200	25
7	120	840	49
12	180	2160	144
16	210	3360	256
20	240	4800	400
12	158	2272	174.8

$$\left.\begin{array}{l}\overline{Y} = a\overline{X} + b \Rightarrow 158 = 12a + b \\ \overline{XY} = a\overline{X^2} + b\overline{X} \Rightarrow 2272 = 174.8a + 12b\end{array}\right\} \begin{array}{l}a = 12.21 \\ b = 11.51\end{array}$$

$$Y = 12.21(x) + 11.51$$

(i) After 10 seconds = $Y = 122.1 + 11.51 = \underline{133.61}$

(ii) Difference in 5 seconds = $Y_0 = 12.21(0) + 11.51 = 11.51$

$Y_5 = 12.21(5) + 11.51 = 72.56$

$\text{diff} = Y_5 - Y_0 = \underline{\underline{61.05}}$

Consider a dataset $D_1 = \{(x^{(1)}, y^{(1)}), \ldots, (x^{(N)}, y^{(N)})\}$ with a linear regression model $y = w_1 x + b_1$ that minimizes the mean squared error.
A second dataset D_2 is created by transforming D_1 as follows:

$$D_2 = \{(x^{(1)} + \alpha, y^{(1)} + \beta), \ldots, (x^{(N)} + \alpha, y^{(N)} + \beta)\}$$

where $\alpha, \beta > 0$ and $w_1 \alpha \neq \beta$.
A new model $y = w_2 x + b_2$ is fitted to D_2. Which of the following statements correctly describes the relationship between the parameters of the models trained on D_1 and D_2?

A. $w_1 = w_2, b_1 = b_2$
B. $w_1 \neq w_2, b_1 = b_2$
C. $w_1 = w_2, b_1 \neq b_2$
D. $w_1 \neq w_2, b_1 \neq b_2$

X_1	Y_1	$X_1 Y_1$	X^2
1	1	1	1
2	2	4	4
3	3	9	9
4	4	16	16
5	5	25	25
3	3	11	11

$3 = 3a + b$
$11 = 11a + 3b$

$a = 1, \; b = 0$

X_2	Y_2	$X_2 Y_2$	X_2^2
2	0	0	4
3	1	3	9
4	2	8	16
5	3	15	25
6	4	24	36
4	2	10	18

$2 = 4a + b$
$10 = 18a + 4b$

$a = 1, \quad b = -2$

$w_1 = w_2, \; b_1 \neq b_2$ (C) Ans.

Question

Q3-2: Which of the following statement is true about outliers in Linear regression?

1. Linear regression is sensitive to outliers
2. Linear regression is NOT sensitive to outliers
3. Can't say
4. None of these

(1) Ans.

Question:
Given the data **{(-1, 1), (2, -5), (3, 5)}** of the form **(x, y)**, we fit a model **y = wx** using linear least-squares regression. The optimal value of **w** is ________.

(Round off to **three decimal places**)

$$y = \omega x$$

Note: Use the second equation of closed form equation without intercept.

$$\overline{XY} = \omega \overline{X^2}$$

X	Y	XY	X^2
-1	1	-1	1
2	-5	-10	4
3	5	15	9
		4/3	14/3

$$\frac{4}{3} = \frac{14}{3}\omega$$

$$\omega = \frac{2}{7} = 0.286$$