Unit-3

Regression (Relation b/w the variables)

→ Regression is a statistical method that helps us understand and predict the relationship between variables.

→ Dependent — Independent.

Dependent ↓ predict.

Independent ↓ जिसके Basis पर करता है।

→ Describes how one variable (dependent variable) changes as another variable (Independent variable) changes.

→ Dependent Variable :- we are trying to predict or explain (Y).

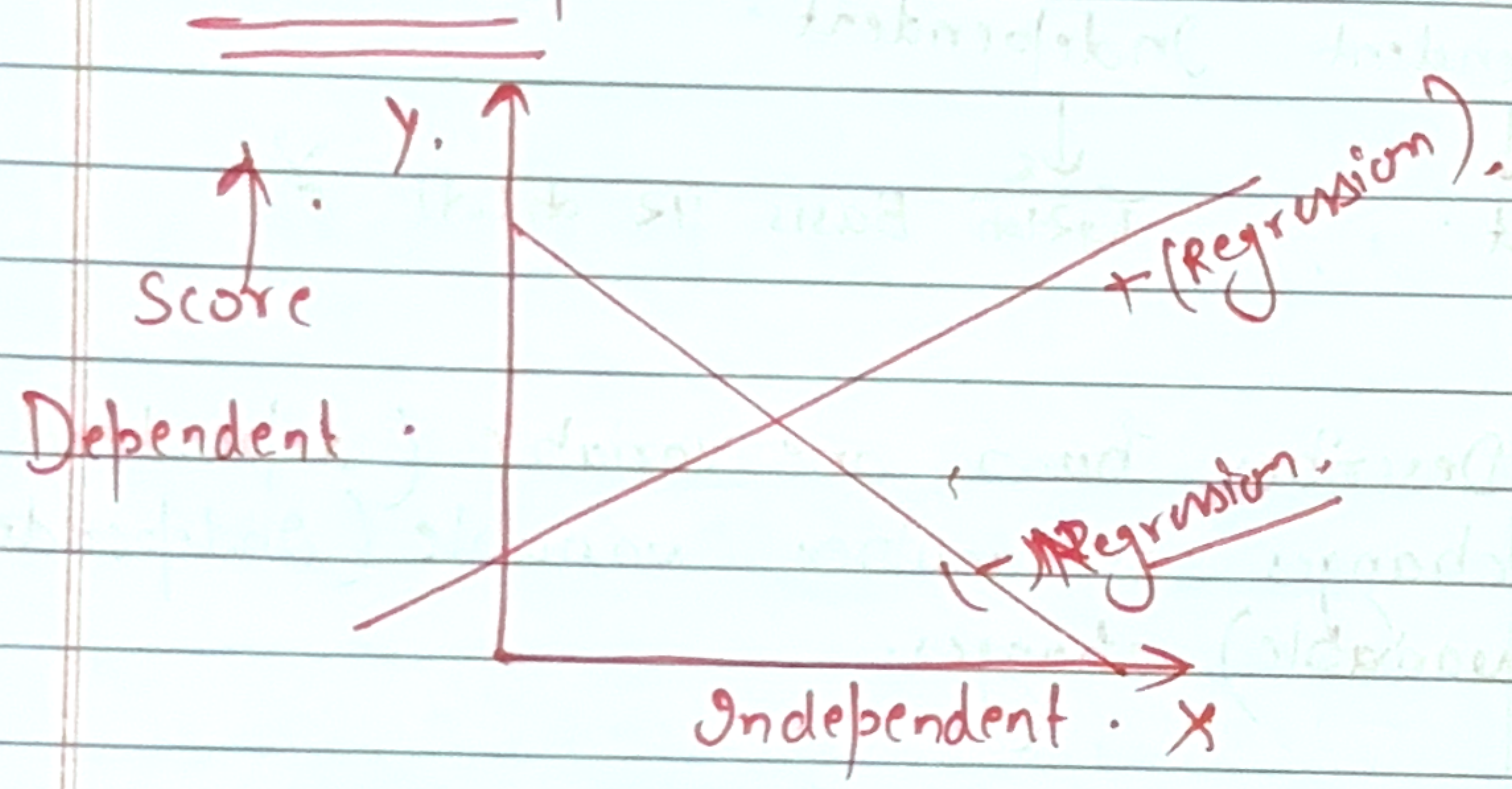
→ Independent Variable :- That are use to predict or explain the changes in the dependent variable (X).

For Example :- Predicting salary based on years of experience, predicting exam score based on study hours, (predicting resale

↓ Dependent

price based on vehicle age.
↓
Independent.

Relationship.



| Age ↑ | Resale ↓. |

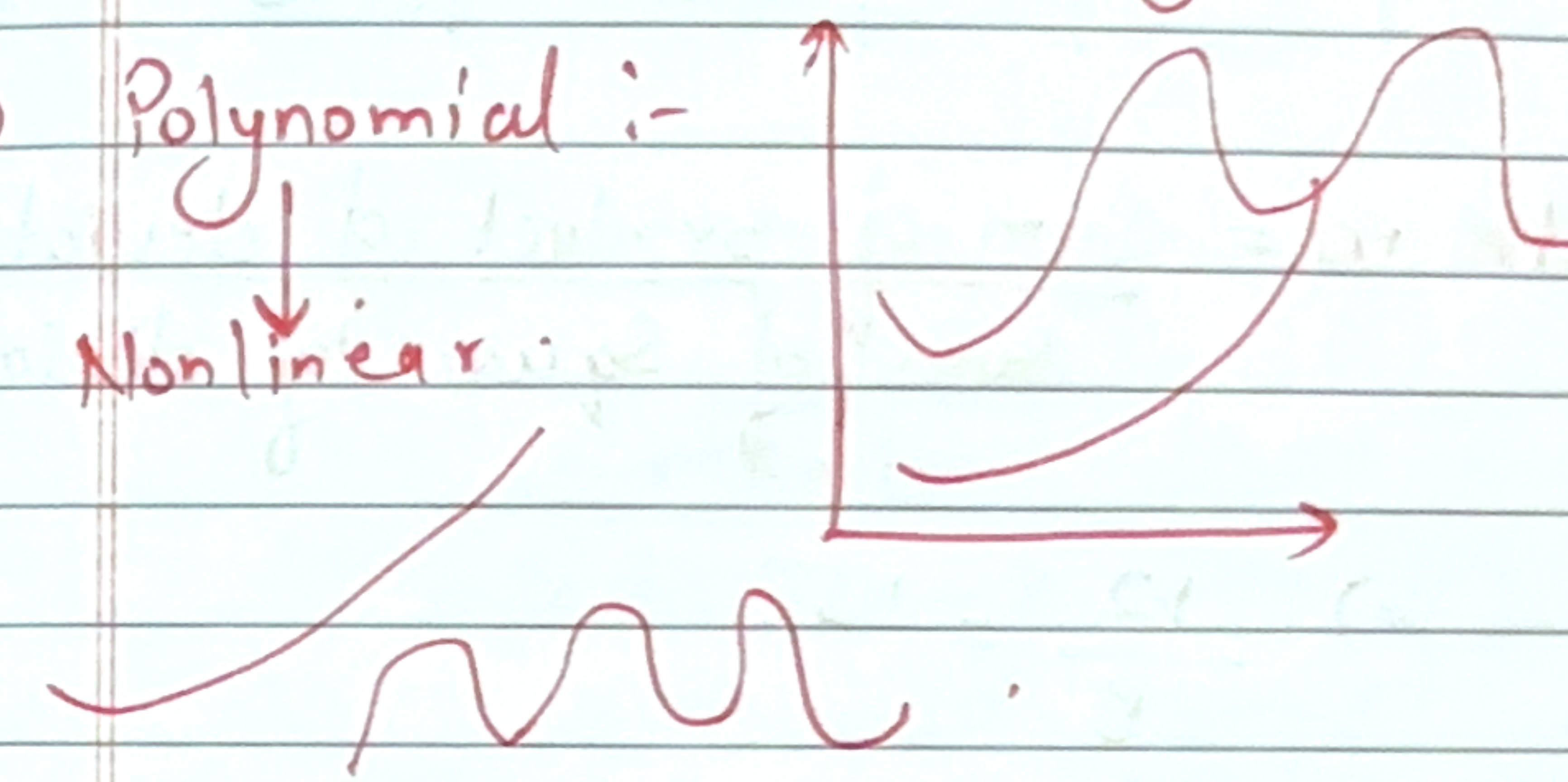
① Linear Regression :- ① Dependent & ① Independent.

→ Relationship between two variables is 1 dependent & 1 Independent then it is Known as Linear Regression.

(2) Multi-Linear Regression.

→ ① Dependent & multiple Independent. then it is Known as Multi linear Regression.

(3) Polynomial :-
↓
Nonlinear.



Equation of Linear Regression.

$\boxed{y = mx + b.}$ → Independent variable.
↓
Dependent variable.

m = slope of the line. (how much y changes for a unit change in x.)
b = Intercept (the value of y when x=0).

Project :- Predicting Pizza prices.
Step-1. Data collection.

Diameter in Inches (x)	Price (y) in Dollars	mean (x)	mean (y)	Deviations (x)	Deviations (y)	Product of deviations	Sum of Product of Deviations	Square of deviations for x.
8	10	10	13	+2	+3	6	12	4
10	13			0	0	0		0
12.	16			-2	-3.	6		4.

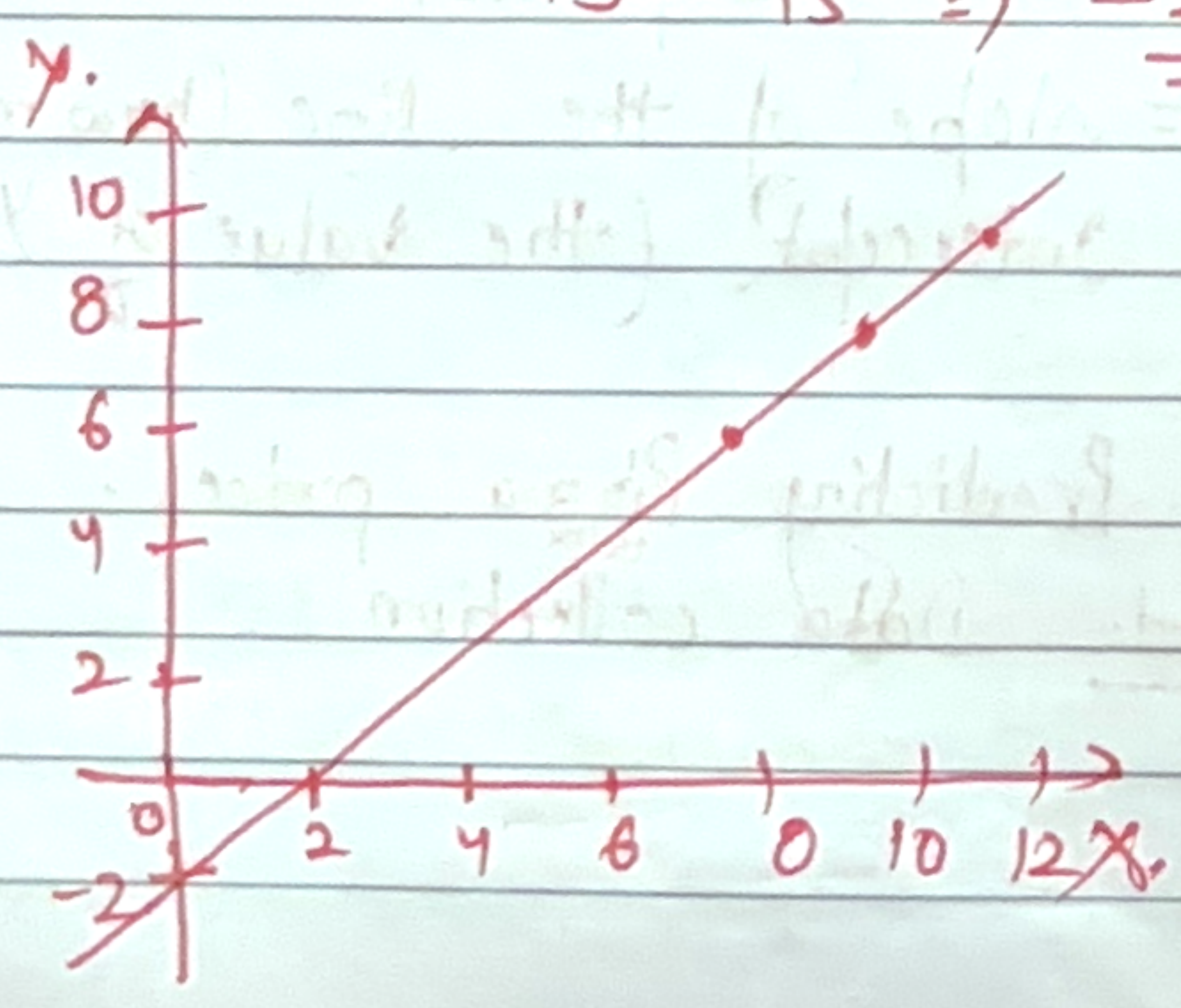
Calculate $m = \dfrac{\text{Sum of product of deviations}}{\text{Sum of Square of deviation for } x}$

$\Rightarrow \dfrac{12}{8} = 1.5$

Calculate $b = \text{Mean of } y - (m * \text{mean of } x)$.

$13 - (1.5 \times 10)$.

$13 - 15 \Rightarrow \underline{\underline{-2}}$.



Pizza → 20 inch. so predict the Dollar.

$y = mx + b$.

$= 1.5 \times 20 - 2$.

$\Rightarrow 30 - 2 = \boxed{28 = y}$.

Mean Square Error [mse].

→ We predict that our model is predicted properly or not. So, we calculate the difference.

→ It is a common metric used to ~~calculate~~ evaluate the accuracy of a linear regression method. It measures the average squared difference between the actual values (observed data) and the predicted value (model's prediction)

→ The goal in linear Regression is to minimize the MSE, as a lower MSE indicates that the model's predictions are closer to the actual values.

$$\frac{1}{n}\sum_{i=1}^{n}(\underbrace{y_i}_{\text{Actual}} - \overbrace{\hat{y}_i}^{\text{predict}})^2.$$

Example.

X (Hours Studied) ↑ Independent	Y (Actual Score)	$\hat{y}$ (Predicted Score) ↑ Predicted value (Dependent)	$(y_i-\hat{y}_i)^2$
1	2	2.5	0.25
2	4	4	0
3	5	5.5	0.25
4	4.5	7	6.25
5	7	8.5	2.25
			Sum = 9

$Y = mX + b$..

$m = ?$

$b = ?$

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

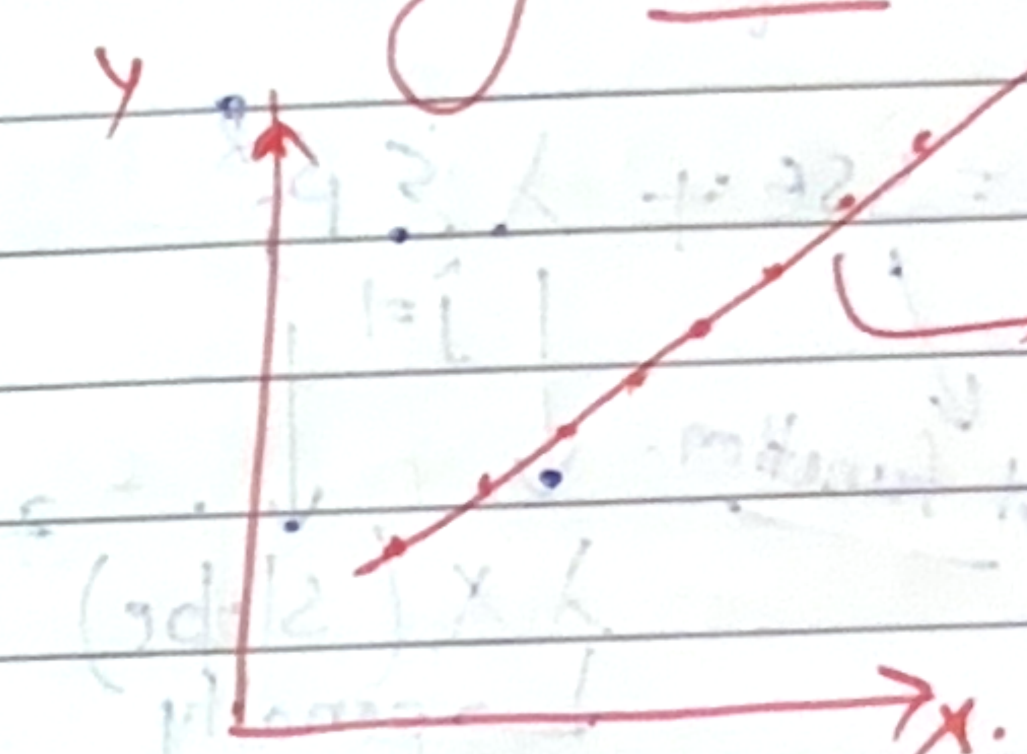
$\frac{9}{n} = \frac{9}{5}$ $= 1.8 \rightarrow$ MSE.

we should reduce this mse then our model becomes more accurate.

$\boxed{MSE = 0.}$

↓

Overfitting Case.



→ Actual & Predicted Value all lie in the same line.

Regularization

→ Ridge Regression (L2) | Lasso Regression (L_1).

→ Ridge Regression and Lasso Regression are both regularization techniques used in linear Regression to prevent overfitting by adding a penalty term to the cost function.

→ Ridge Regression (L2 Regularization) :- Adds a

Cost function = $(\text{Prediction} - \text{Actual})^2$

↓

Sum of Squared Error.

penalty proportional to the sum of the squares of the coefficients.

$$CF = SSE + \lambda \sum_{j=1}^{p} \beta_j^2$$

Cost function. $\lambda \times (\text{slope})^2$ → penalty.

(2) Lasso Regression (L1 Regularization) :- Adds a penalty proportional to the sum of the absolute values of the coefficients.

$$SSE + \lambda \sum_{j=1}^{p} |\beta|.$$

$\lambda \times |\text{slope}|$.

CF ↓

Example.

X	Y. (Actual)	Prediction	$(\text{Prediction-Actual})^2$
1	2	2	0
2	4	4	0
3	6	6	0
4	8.	8	0

Now Apply Simple Linear Regression.

$Y = \beta_0 + \beta_1 X.$

Intercept, slope.

Or

$\boxed{Y = MX + C.}$

$\beta_0 = 0.$

$\beta_1 = 2.$

$Y = 0 + 2X.$

$\boxed{Y = 2X.}$

with the help of formula.

Error = 0 (SSE).

→ Overfitting → Training 95%, → Testing 65%.

$L2 = SSE + \lambda \sum_{i=1}^{p} \beta_j^2.$

$0 + \lambda [2]^2 \Rightarrow L2 = 4\lambda$ → Let $\lambda = 1$ → $\boxed{L2 = 4}$

$L_1 = SSE + \lambda|slope|$.

$L_1 = 0 + 1 \times 2$.

$\boxed{L_1 = 2}$.

It add some penalty and reduce the problem of overfitting.

Ex-2.

X	Y	$\hat{y}$ (predicted)	$(y_i - \hat{y}_i)^2$
1	2	2.8	0.64
2	4	3.4	0.36
3	5	4	1
4	4	4.6	0.36
5	5	5.2	0.04
			2.4

$y = mx + c$. → $m = 0.6$, $c = 2.2$ } with the help of linear Regression formula that I have explained earlier.

for finding the predicted values.

put $x = 1$,

$y = 0.6 \times 1 + 2.2$.

$y = 2.8$

put $x = 2$.

$y = 0.6 \times 2 + 2.2$.

$y = 3.4$.

put $x = 3$. $y = mx + c$.

$y = 0.6 \times 3 + 2.2$.

$y = 4$.

Put $x = 4$. $y = mx + c$.

$y = 0.6 \times 4 + 2.2$.

$y = 4.6$.

$x = 5$ $y = mx + c$.

$y = 0.6 \times 5 + 2.2$.

$y = 5.2$.

$\boxed{SSE = 2.4}$.

$L_2 = SSE + \lambda \times (slope)^2$.

$2.4 + \lambda \times (0.6)^2$. $\lambda = 1$ (Let)

$2.4 + 1 \times (0.6)^2$.

$\boxed{L_2 = 2.76}$.

$L_1 = SSE + \lambda \times |slope|$.

$2.4 + 1 \times 0.6$.

$2.4 + 0.6$

$\boxed{L_1 = 3}$.

Bias & Variance

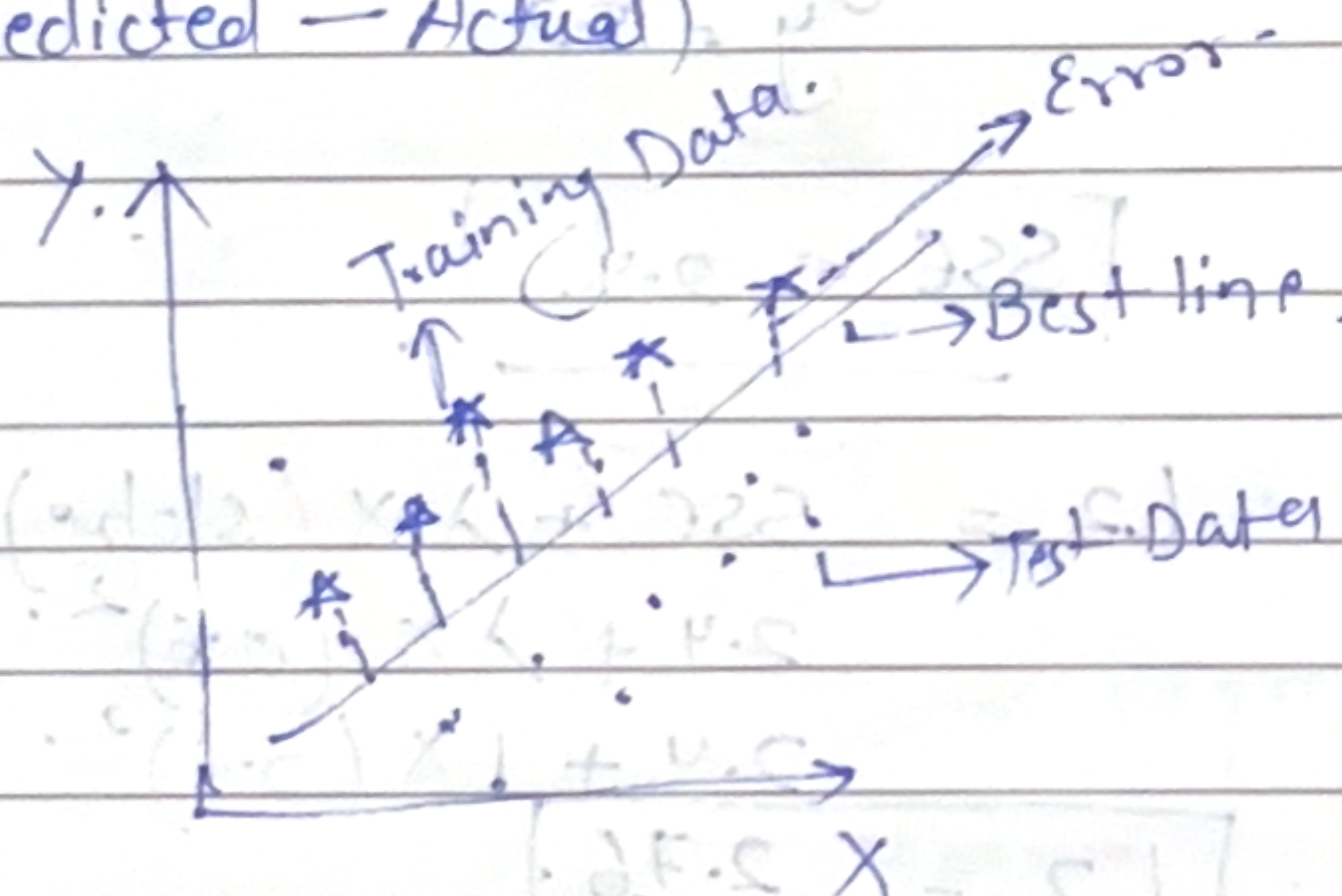
Data ⟶ MLM (Machine Learning Model) ⟶ o/p (predicted output)

1000 (Data) ⟶ 700 (Training), 300 (Testing)

⟶ Actual o/p

We find that the error comes or not.

Error = (Predicted − Actual)



If the training data having more Error then it means High Bias.

Train Data ↑ } (High Bias)
Test Data ↑

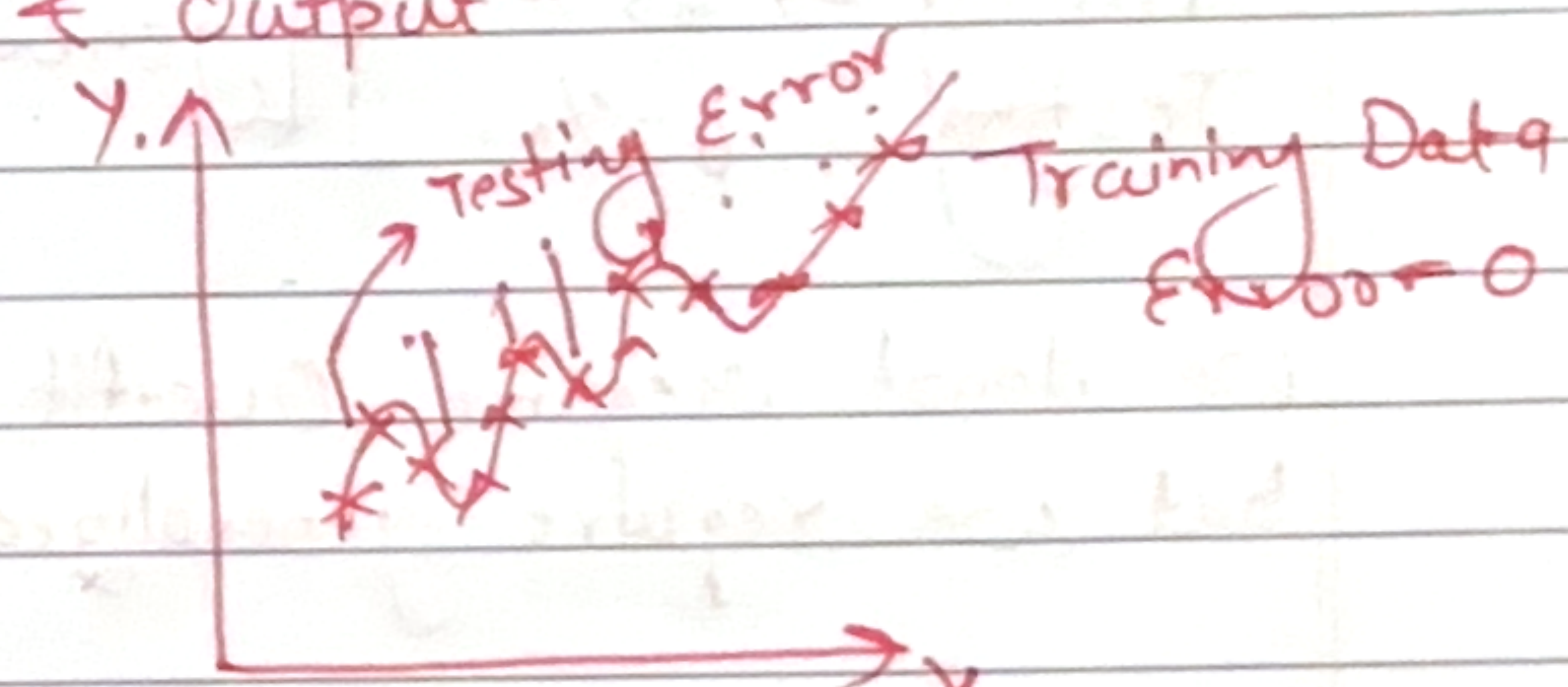
Underfit - (If any model is not working properly at the time of training).

→ Bias Refers to Errors due to overly simplistic assumptions in the learning algorithm.

→ Bias always deals with Training Data.

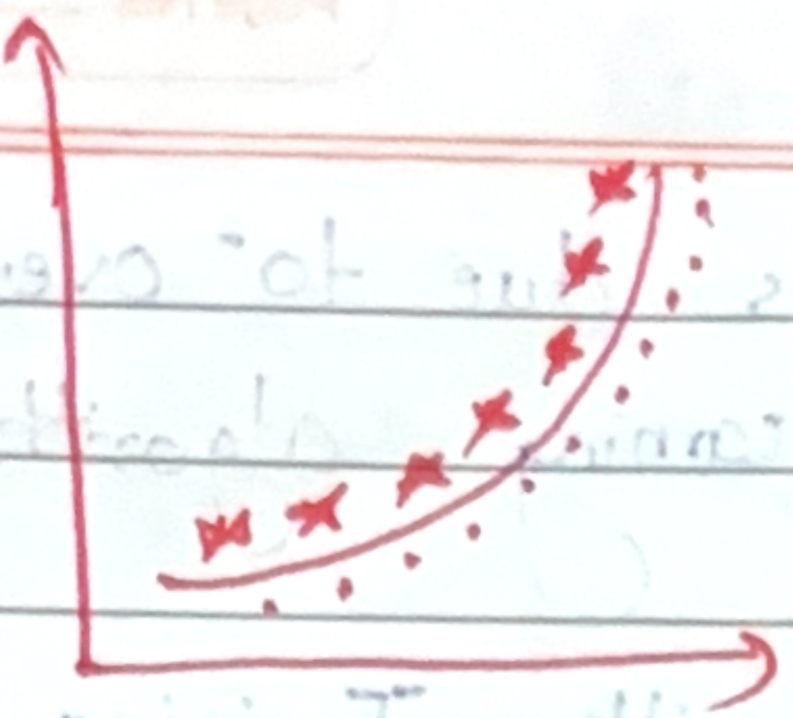
→ High Bias :- When the model is too simple, it underfits the data. This means the model does not capture important patterns in the data, leading to poor performance on both the training & testing sets.

→ Low Bias :- When the model is sufficiently complex and can capture the true relationship b/w Input & Output.



Low Bias.
High Variance (Overfitting)

Date:
Page No:



Low Bias. [generalized model].
Low Variance.

Training → 90% ↑ } overfitting X
Testing → 20% ↓ }

Training → 20% ↓. [High variance]. X
Testing → 90% ↑ [Underfitting].

Training → 60%.] [generalized model]
Testing → 62%.]

We donot require Overfit & Underfit model but we require generalized model.