Linear Regression in Machine Learning

→ **Regression**: A statistical technique that relates a dependent variable to one or more independent variables.

→ **AI** (Artificial Intelligence)

↓

ML: Machine Learning is a branch of AI that focuses on the development of algorithms and statistical models that can learn from and make predictions on data.

↓

Linear Regression: Type of ML algorithm learns from Supervised dataset (labelled data) to predict dependent variable based on independent variables.

[Independent Variables/Features —Linear Relationship→ Dependent Variable/feature]

Linear Regression (L.R)

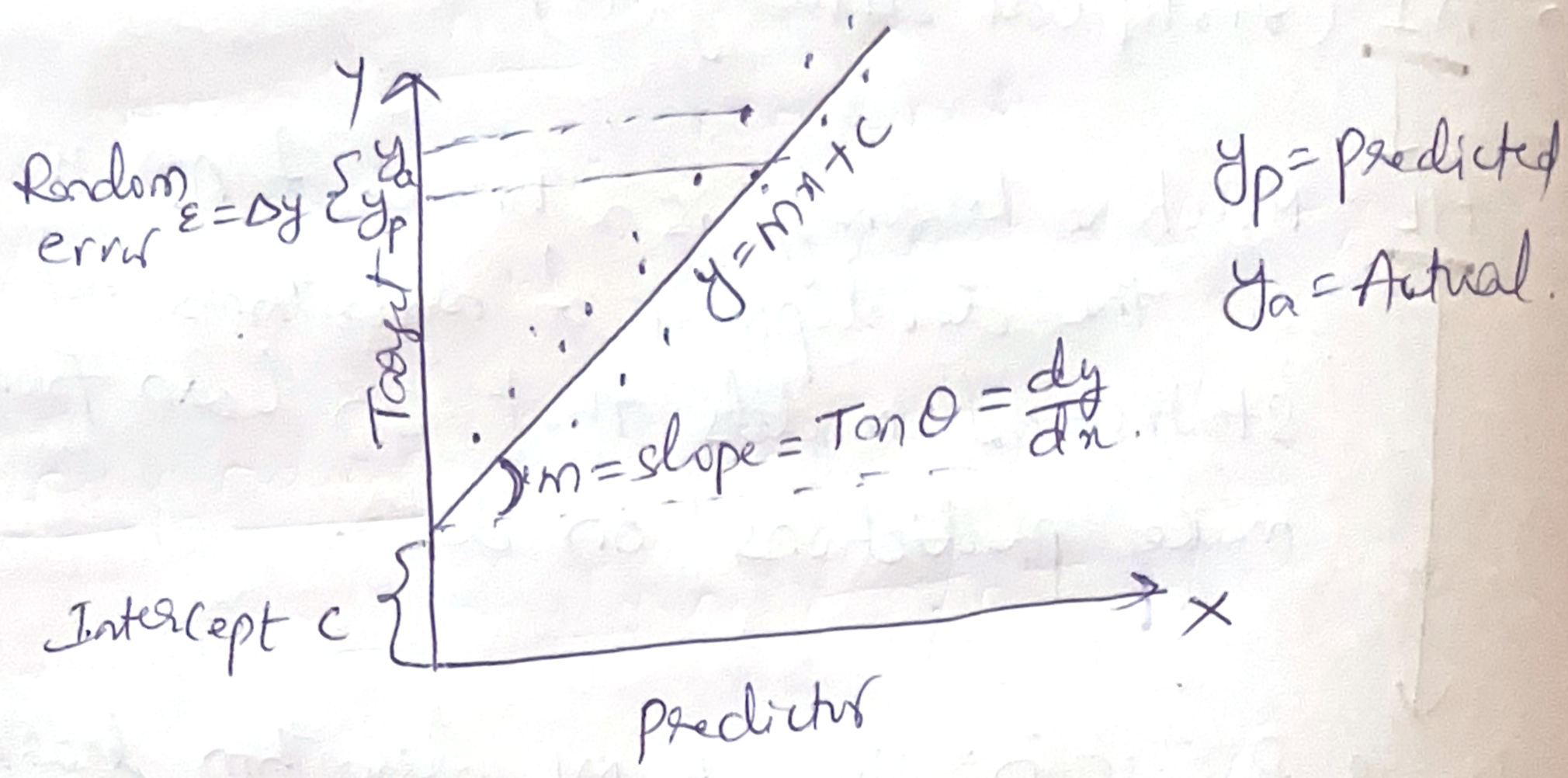
- **Simple L.R** (One Independent feature) $\boxed{y = mx + c}$
- **Multiple L.R** (>1 Independent features) $\boxed{y = m_1x_1 + m_2x_2 + \cdots + c}$
- **Univariate L.R** (One dependent variable)
- **Multivariate** (>1 dependent variable)

Goal: To find best fit Line/Equation to predict 'y' by 'x'.

→ Least error in best fit line.

→ error = $y_{predict} - y_{actual}$

→ y = dependent or Target
x = Independent or Predictor.



Eg: x = experience
y_a = salary. actual ; y_p = predicted
θ_1 = Intercept
θ_2 = slope / coe. of x.
y_a = actual
$i = 0, 1, 2, \ldots$

$$\boxed{y_p = \theta_1 + \theta_2 x}$$

$$\boxed{\text{Error: } \varepsilon = y_{p_i} - y_{a_i}}$$
Residual

Cost function / Loss function = error. (J)

In L.I Mean Squared Error (MSE) is used.
(J)

$$\boxed{J = \frac{1}{n} \sum_{i=0}^{n} \varepsilon_i = \frac{1}{n} \sum_{i=0}^{n} (y_{p_i} - y_{a_i})}$$

→ How to minimise Cost function?

Ans: Using Gradient Descent.
Gradually. → decreasing
(Real meaning is slope)

gradient descent using derivatives.

$$J = \Sigma (y_p - y_a)^2 \times \frac{1}{n}$$

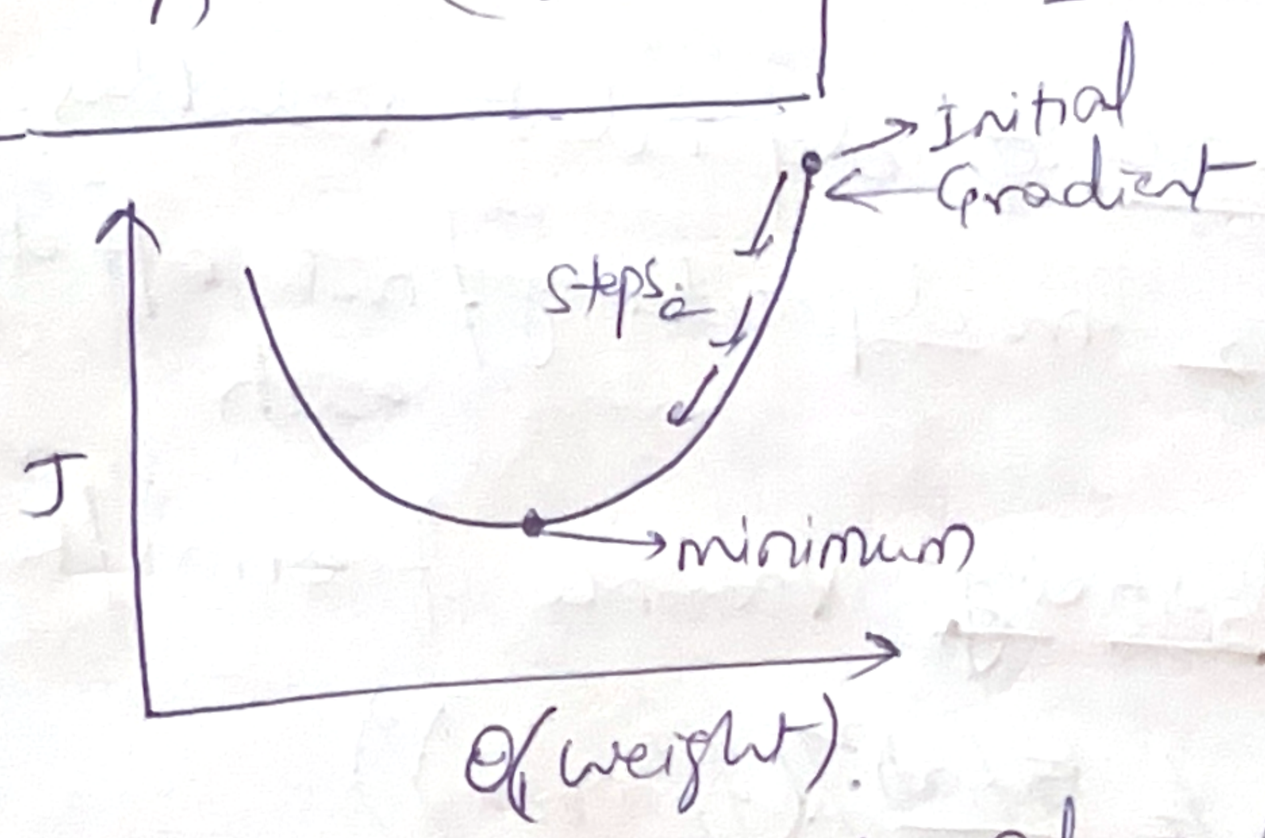
$$= \frac{1}{n} \Sigma (\Delta y_i)^2 = \frac{1}{n} \Sigma (y_p - y_a)^2 \text{ or } \frac{1}{n} \Sigma (y_a - y_p)^2$$

Same (square).

$$J = \frac{1}{n} \Sigma (\theta_1 + \theta_2 x - y_a)^2$$

$$\boxed{\frac{\partial J}{\partial \theta_1} = \frac{2}{n} \Sigma (\Delta y_i)} = J'_{\theta_1}$$

$$\boxed{\frac{\partial J}{\partial \theta_2} = \frac{2}{n} \Sigma (\Delta y_i) \cdot x_i} = J'_{\theta_2}$$



By changing/placing different values of θ_1 & θ_2, we get minimum cost/loss/J.

Learning rate (α) :- The changing of θ_1 & θ_2.
Improving Learning.

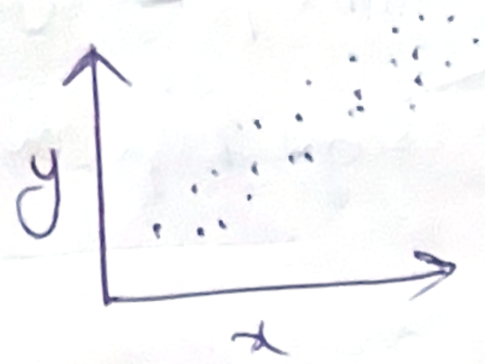
$$\theta_1 = \theta_1 - \alpha J'_{\theta_1}$$

$$\theta_2 = \theta_2 - \alpha J'_{\theta_2}$$

Assumptions / Requirements :

→ Assumptions / Requirements Simple Linear Regression

1. Linearity: Linear relationship b't y & x.



Linear

Non-Linear

(Simple terms, we can make best fit equation) or not.

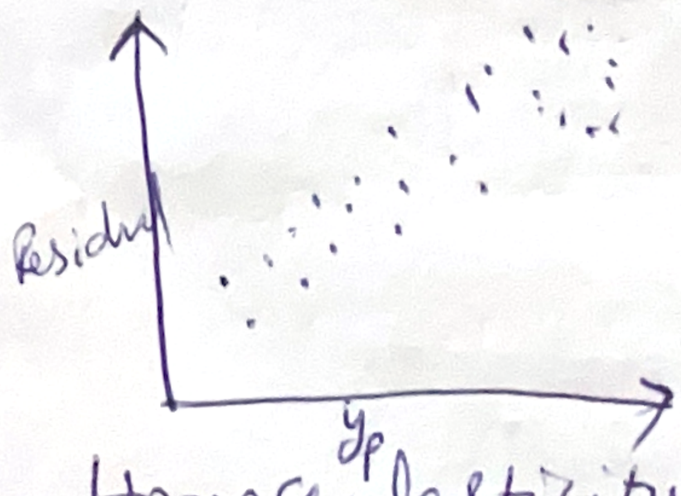
2. Independence of dependent variables.

eg. x_1, x_2 no dependency / not affecting values of each

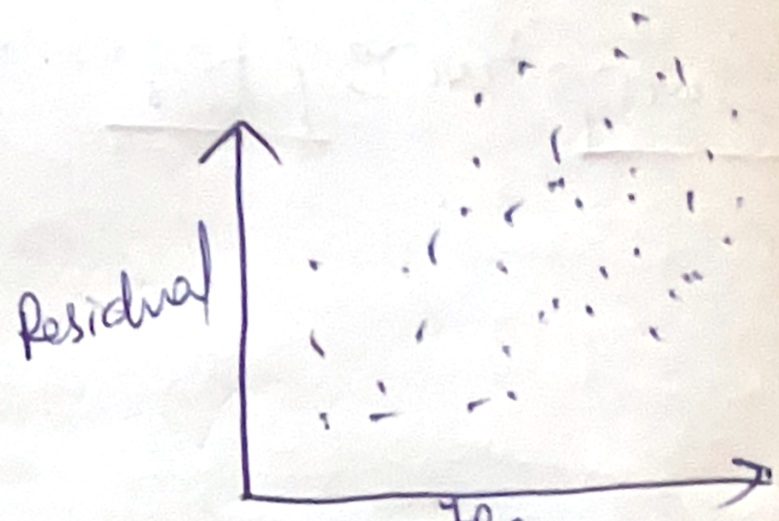
3. Homoscedasticity: variance of errors / Residuals Constant.

(variance $= \frac{(\bar{x}-x)^2}{n}$ (formula))

Across all levels of independent variable(s), the variance of errors / Residuals is Constant. This indicates that the amount of independent variable(s) has no impact on the variance of errors. If the variance of residuals is not Constant, then L.I will not be an accurate model.

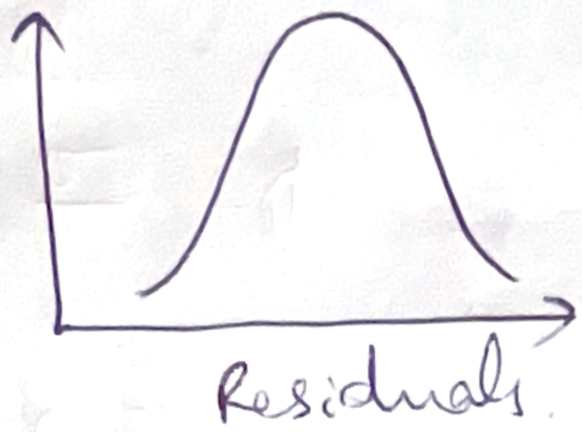


Homoscedasticity
(Same variance)



Heteroscedasticity

4. Normality: The residuals should be normally distributed.



→ Assumptions of multiple Linear Regression:

1. No multicollinearity: No high correlation b/t the independent variables.

2. Additivity: The model assumes that the effect of changes in a predictor variable on the response variable is consistent regardless of the values of the other variables. This assumption implies that there is no interaction b/t variables in their effects on the dependent variable.

3. Feature Selection:
Carefully selecting independent variables.

4. Overfitting: This occurs when the model fits the training data too closely, capturing noise or random fluctuations.

→ Multicollinearity: It occurs when independent variables in a multiple linear model are highly correlated.

Detecting:

1) Correlation Matrix: Among independent variables.

High Correlation : 1 (+ve)
−1 (−ve)

No Correlation : 0

(2) VIF (Variance Inflation Factor): VIF is a measure that quantifies how much the variance of an estimated regression coefficient (y) increases if your predictors (x's) are correlated.

A high VIF (>10) suggests multicollinearity.

→ Evaluation Metrics for Linear Regression:-
(Checking accuracy of model)
Evaluation Measures/Metrics can be used to determine the strength of any linear Regression model. These assessment metrics often give an indication of how well the model is producing the observed outputs.

The most common measures are:

(1) MSE (2) MAE (3) RMSE (4) R-squared.
(5) Adjusted R-squared error.

① Mean Squared Error (MSE):

$$\boxed{MSE = \frac{1}{n} \sum (y_{a_i} - y_{p_i})^2}$$

n = No. of data points.
y_{a_i} = y-actual at ith point
y_{p_i} = y-predicted "

MSE is sensitive to outliers as large errors contributes significantly to the overall score.

② Mean Absolute Error (MAE):

$$\boxed{MAE = \frac{1}{n} \sum |y_{a_i} - y_{p_i}|}$$

Lower MAE value indicates better model performance.
Not sensitive to outliers as we consider absolute differences.

③ Root Mean Squared Error (RMSE):
(or) Residual standard Error (RSE).

$$\boxed{RMSE = \sqrt{\frac{RSS}{n}} = \sqrt{\frac{\sum (y_{a_i} - y_{p_i})^2}{n}}}$$

RSS = Residual Sum of squares.

RMSE is not as good of a metric as R-squared.
RMSE can fluctuate when the units of the variables vary since its value is dependent on the variable's unit (it is not a normalized measure).

Coe. of determination (R-squared) R^2

R-squared is a statistic that indicates how much variation the developed model can explain.

Range: 0 to 1.

Greater R-squared, better model.

$$R^2 = 1 - \left(\frac{RSS}{TSS}\right)$$

RSS = Residual Sum of Squares

$$= \sum (y_{a_i} - y_{p_i})^2$$

TSS = Total Sum of Squares

$$= \sum (y_{a_i} - \bar{y}_{a_i})^2$$

$$R^2 = 1 - \frac{\sum (y_{a_i} - y_{p_i})^2}{\sum (y_{a_i} - \bar{y}_{a_i})^2}$$

$\bar{y}_{a_i}$ = y actual mean.

R-squared metric is a measure of the proportion of variance in the dependent variable that is explained the independent variables in the model.

Adjusted R-squared Error:

Adjusted R-squared (R^2) measures the proportion of variance in the dependent variable that is explained by independent variables in a regression model.

Adjusted R^2 accounts the no. of predictors in the model and penalizes the model for including irrelevant predictors that don't contribute significantly to explain the variance in the dependent variables.

$$\boxed{\text{Adjusted } R^2 = 1 - \frac{(1-R^2)(n-1)}{n-k-1}}$$

n = no. of observations.
k = no. of predictors / features.
R^2 = Coe. of determination.

Adjusted R^2 helps to prevent overfitting. It penalizes the model with additional predictors that do not contribute significantly to explain the variance in the dependent variable.

→ Regularization : Preventing Overfitting.
λ = Overall regularization strength.

(1) Lasso Regression (L1 Regularization)

It adds a penalty term to the linear regression objective function to prevent overfitting.

The objective function after applying Lasso.

$$J(\theta) = \underbrace{\frac{1}{2m}\sum_{i=1}^{m}(y_{p_i} - y_a)}_{\text{Least Squares loss}} + \underbrace{\lambda \sum_{j=1}^{n} |\theta_j|}_{\text{L1 regularization term which penalizes the sum of absolute values of the regression coe. } \theta_j}$$

(2) Ridge Regression (L2 Regularization): Prevent overfitting by penalizing large coefficient in L.R. Useful in case of multicollinearity where predictors are highly co-related.

$$\boxed{J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(y_{p_i} - y_{a_i}) + \lambda \sum_{j=1}^{n} \theta_j^2}$$

→ Second term = penalizes the sum of squares of θ_j

(3) Elastic Net Regression: Hybrid regularization.
(L1 + L2)

$$\boxed{J(\theta) = \underbrace{\frac{1}{2m}\sum_{i=1}^{m}(y_{p_i} - y_{a_i})}_{\text{Least Squares Loss}} + \underbrace{\alpha\lambda\sum_{j=1}^{n}|\theta_j|}_{L1} + \underbrace{\frac{1}{2}(1-\alpha)\lambda\sum_{j=1}^{n}\theta_j^2}_{L2}}$$

λ = Overall regularization strength.
α = Controls the mix b't L1 & L2 regularization.

How do we get the coefficients of Simple Linear Regression:

Eg: Finding the regression equation using Least Squares method.

X	Y
5	1
4	2
3	3
2	4
1	5

Find the linear regression equation.

Sol^n: Straight line equation: $\boxed{y = a + bx}$ — (1)

$\Rightarrow \Sigma y = \Sigma(a + bx)$

$\boxed{\Sigma y = an + b\Sigma x}$ — (2) $\quad \because \Sigma 1 = n.$

(1) $* \Sigma x \Rightarrow \boxed{\Sigma xy = a\Sigma x + b\Sigma x^2}$ — (3)

② & ③ are called normal equations.

x	y	x^2	$x * y$
5	1	25	5
4	2	16	8
3	3	9	9
2	4	4	8
1	5	1	5
$\Sigma x = 15$	$\Sigma y = 15$	$\Sigma x^2 = 55$	$\Sigma xy = 35$

& $n = 5$ observations.

② $\Rightarrow$ $5a + 15b = 15$

$\Rightarrow$ $\boxed{a + 3b = 3}$ —— ④

$3a + 9b = 9$
$3a + 11b = 7$
$-2b = 2$
$b = -1$.
$a = 6$.

③ $\Rightarrow$ $15a + 55b = 35$

$\Rightarrow$ $\boxed{3a + 11b = 7}$ —— ⑤

$\therefore$ $\boxed{b = -1 \ \& \ a = 6}$

$\therefore$ The regression equation is

$$\boxed{y = 6 - x}$$