Simple Linear Regression

- **Simple Linear Regression (1 feature, 1 output)** में **w (slope)** और **b (bias/intercept)** कैसे निकालते हैं।
  मैं एक **छोटा dataset** लूंगा और formulas के साथ समझाऊँगा।
- Y = wx+b is the formula of Simple Linear Regression.

Step 0: Small Dataset

X	Y
1	2
2	3
3	5

हम model मानते हैं:

$$y_{predicted} = wx + b$$

हमको निकालना है **w** और **b**।

What is w?

- **W is slope/coefficient/weight**
- **Formula for find the w-**

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

$$b = \bar{y} - w\bar{x}$$

जहाँ:

- $\bar{x}$ = X का mean
- $\bar{y}$ = Y का mean

Finding Mean of X, Y

- **Find the Mean**

Means निकालो

$$\bar{x} = \frac{1+2+3}{3} = 2$$

$$\bar{y} = \frac{2+3+5}{3} = 3.33$$

Finding numerator of w

Numerator: $\sum(x_i - \bar{x})(y_i - \bar{y})$

x_i	y_i	x_i - x̄	y_i - ȳ	(x_i-x̄)*(y_i-ȳ)
1	2	1-2 = -1	2-3.33=-1.33	(-1)*(-1.33)=1.33
2	3	2-2 = 0	3-3.33=-0.33	0*-0.33=0
3	5	3-2=1	5-3.33=1.67	1*1.67=1.67

Sum = 1.33 + 0 + 1.67 = 3

Find Denominator of w

Denominator: $\sum(x_i - \bar{x})^2$

x_i	x_i - x̄	(x_i - x̄)^2
1	-1	1
2	0	0
3	1	1

Sum = 1 + 0 + 1 = 2

Finally get the w and b

w निकालो

$$w = \frac{3}{2} = 1.5$$

---

b निकालो

$$b = \bar{y} - w\bar{x} = 3.33 - 1.5 * 2 = 3.33 - 3 = 0.33$$

Model तैयार

$$y_{predicted} = 1.5x + 0.33$$

Predicted Value and Error

Predicted Values और Error

X	Y_actual	Y_predicted	Error = Y_actual - Y_predicted
1	2	1.5+0.33=1.83	0.17
2	3	3+0.33=3.33	-0.33
3	5	4.5+0.33=4.83	0.17

Final Summary

- **Summary**

Parameter	Value
w (slope)	1.5
b (bias)	0.33
y_predicted	1.5x + 0.33

Errors and Evaluation metrics

- 1. RSS (Residual Sum of Squares)

RSS = (Error)2

Formula:

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

- y_i = actual value
- ŷ_i = predicted value

Meaning:

> Model की predictions और actual values के बीच **total squared error** I
> जितना छोटा RSS होगा, model उतना अच्छा fit हुआ।

Errors and Evaluation metrics

2. MSE (Mean Squared Error)

Formula:

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

- Basically RSS को **data points की संख्या** से divide कर दिया।
- Scale independent होता है, जिससे बड़े datasets में comparison आसान होता है।

3. RMSE (Root Mean Squared Error)

Formula:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2}$$

- MSE का square root लेते हैं ताकि **error की unit original y की unit के same** हो जाए।
- Intuitive sense देता है कि average prediction error कितनी unit दूर है।

Errors and Evaluation metrics

4. MAE (Mean Absolute Error)

Formula:

$$MAE = \frac{1}{n}\sum |y_i - \hat{y}_i|$$

- यहाँ **absolute error** लिया जाता है।
- Outliers पर MSE/RMSE की तुलना में कम sensitive है।

Errors and Evaluation metrics

5. R^2 Score (Coefficient of Determination)

Formula:

$$R^2 = 1 - \frac{RSS}{TSS}$$

जहाँ:

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

- TSS = total variance in y
- R^2 = model ने variance में कितना explain किया
- Value 0–1 के बीच (कभी negative भी हो सकती है अगर model बहुत खराब है)
- **1** → perfect fit, **0** → baseline model

Errors and Evaluation metrics

6. Adjusted R^2

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1}$$

- n = data points
- p = number of features
- Useful for multiple regression (multi-feature)
- Penalizes extra features जो मदद नहीं करते

---

7. Other minor metrics

- **Explained Variance Score** – similar to R^2, सिर्फ variance के हिसाब से
- **Mean Squared Log Error (MSLE)** – जब y बहुत skewed हो या log scale में better हो
- **Median Absolute Error** – जब outlier बहुत हो

Summary of error

Quick Summary Table

Metric	Formula	Use-case
RSS	$\Sigma(y-\hat{y})^2$	Total squared error
MSE	RSS/n	Average squared error
RMSE	$\sqrt{}$MSE	Average error in original unit
MAE	Σ	$y-\hat{y}$
R^2	1 - RSS/TSS	Fraction of variance explained
Adjusted R^2	1 - $((1-R^2)(n-1)/(n-p-1))$	R^2 adjusted for #features
MSLE / Median AE	-	Special cases

Use these all error's in sklearn

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
..... Train_test_split.....
model = LinearRegression()
model.fit(X, y)
..... Find the Prediction.....
y_pred = model.predict(X)
print("Predicted values:", y_pred)
..... Coefficients (w) aur bias (b) nikaalo .....
print("Slope (w):", model.coef_[0])
print("Intercept (b):", model.intercept_)
```

Use these all error's in sklearn

- ..... RSS (Residual Sum of Squares) .....

```
RSS = np.sum((y - y_pred)**2)
print("RSS:", RSS)
```

..... MSE (Mean Squared Error) .....

```
MSE = mean_squared_error(y, y_pred)
print("MSE:", MSE)
```

..... RMSE (Root Mean Squared Error) .....

```
RMSE = mean_squared_error(y, y_pred, squared=False)
print("RMSE:", RMSE)
```

Use these all error's in sklearn

..... MAE (Mean Absolute Error) .....

```
MAE = mean_absolute_error(y, y_pred)
print("MAE:", MAE)
```

..... R^2 Score .....

```
R2 = r2_score(y, y_pred)
print("R² Score:", R2)
```