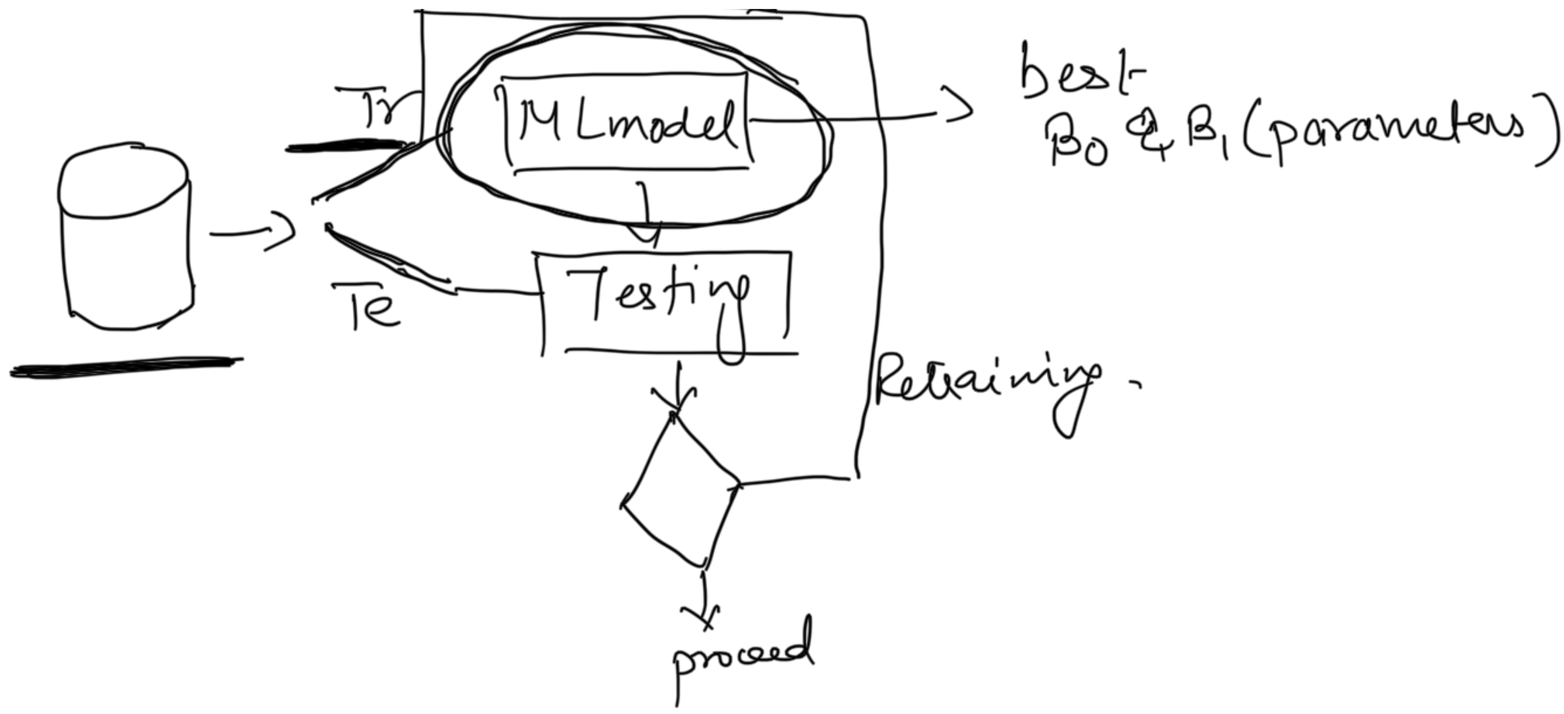


ML model

features & target → $\boxed{ML}$ → logic | pattern.

Classification

target → categorical

sentiments
- negative
- positive
- neutral

Regression.

target → continuous

sales, rainfall.

Regression

features & target → $\boxed{ML}$ → logic

features → X, target → Y → $\boxed{ML}$ → f

such that

$$Y = f(X)$$

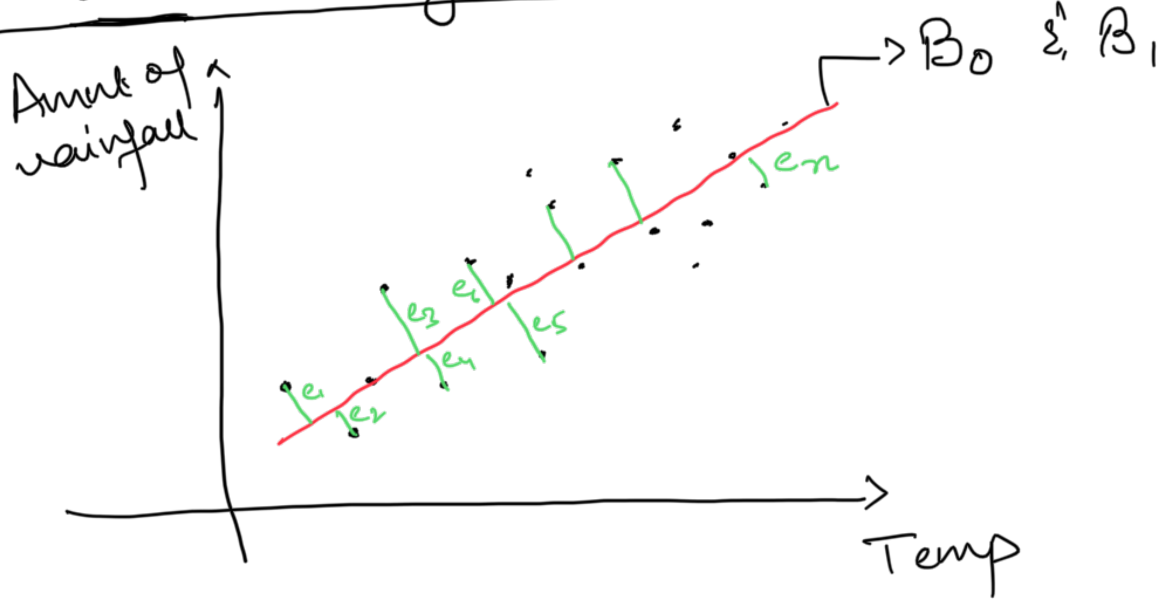
eg predicting price of a house.

IV → carpet area

price = $\boxed{f}$ (carpet area)
↳ red line

The linear models

① Simple Linear Regression.



$$\frac{e_1^2+e_2^2+e_3^2+e_4^2 \dots + e_n^2}{n} = \text{MSE} \downarrow$$

General equation of the line:

$$Y = B_0 + B_1 X \qquad \{y = mx + c\}$$

Best fit line. How?? By minimize the errors.

Draw a line $L_1 \rightarrow E_1 \rightarrow$

$L_2 \rightarrow E_2 < E_1$

$L_3 \rightarrow E_3 < E_2 < E_1$

⋮

$B_0 + B_1 X$

$B_2 + B_3 X$

$B_4 + B_5 X$

⋮

$B_n + B_{n+1} x$

Objective: In the eqn of line

if you want min error then find best B_0 and B_1 (parameters)

$$Y = B_0 + B_1 X$$

parameters / coefficients

$B_0 \rightarrow$ intercept $B_1 \rightarrow$ slope

a. Least square method.

$$Y = B_0 + B_1 X \longrightarrow \text{equation } 1$$

$$Y_i = B_0 + B_1 x_i + e_i \longrightarrow \text{equation } 2$$

Y_i = Actual value

$\hat{Y}_i$ = Predicted value.

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 x_i \quad - \quad 1 \quad \{i = 1, 2, \ldots n\}$$

$$Y_i = \hat{Y}_i + \hat{e}_i \quad - \quad 2 \rightarrow$$

eg $\boxed{S.P}$ you have a model (B_0, B_1)

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 (\text{day})$$

$$= 202$$

$$Y_i = 205$$

$Y_i = 202 + 3$
$\hat{Y}_i + \hat{e}_i$

$\hat{Y}_1 = B_0 + B_1 X_1$; $\hat{Y}_2 = B_0 + B_1 X_2$ | .... $\hat{Y}_n = B_0 + B_1 X_n$

$\underline{Y_1} = \underline{\hat{Y}_1} + \hat{e}_1$; $\underline{Y_2} = \underline{\hat{Y}_2} + \hat{e}_2$ | $\underline{Y_n} = \underline{\hat{Y}_n} + \hat{e}_n$

$$\hat{e}_1^2 + \hat{e}_2^2 + \ldots + \hat{e}_n^2 = \sum_{i=1}^{n} \hat{e}_i^2$$

eg:-

	Actuals	Preds	e
Monday	202	205	$\cancel{-3}$
Tuesday	203	200	$\cancel{3}$

$\Sigma = 0$ 6

$$RSS = \sum_{i=1}^{n} e_i^2$$

$$= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$Y_i = \hat{Y}_i + \hat{e}_i$

$\hat{Y}_i = B_0 + B_1 X_1$

$$RSS = \sum_{i=1}^{n} (Y_i - \hat{B}_0 - \hat{B}_1 X_1)$$

Minimize RSS

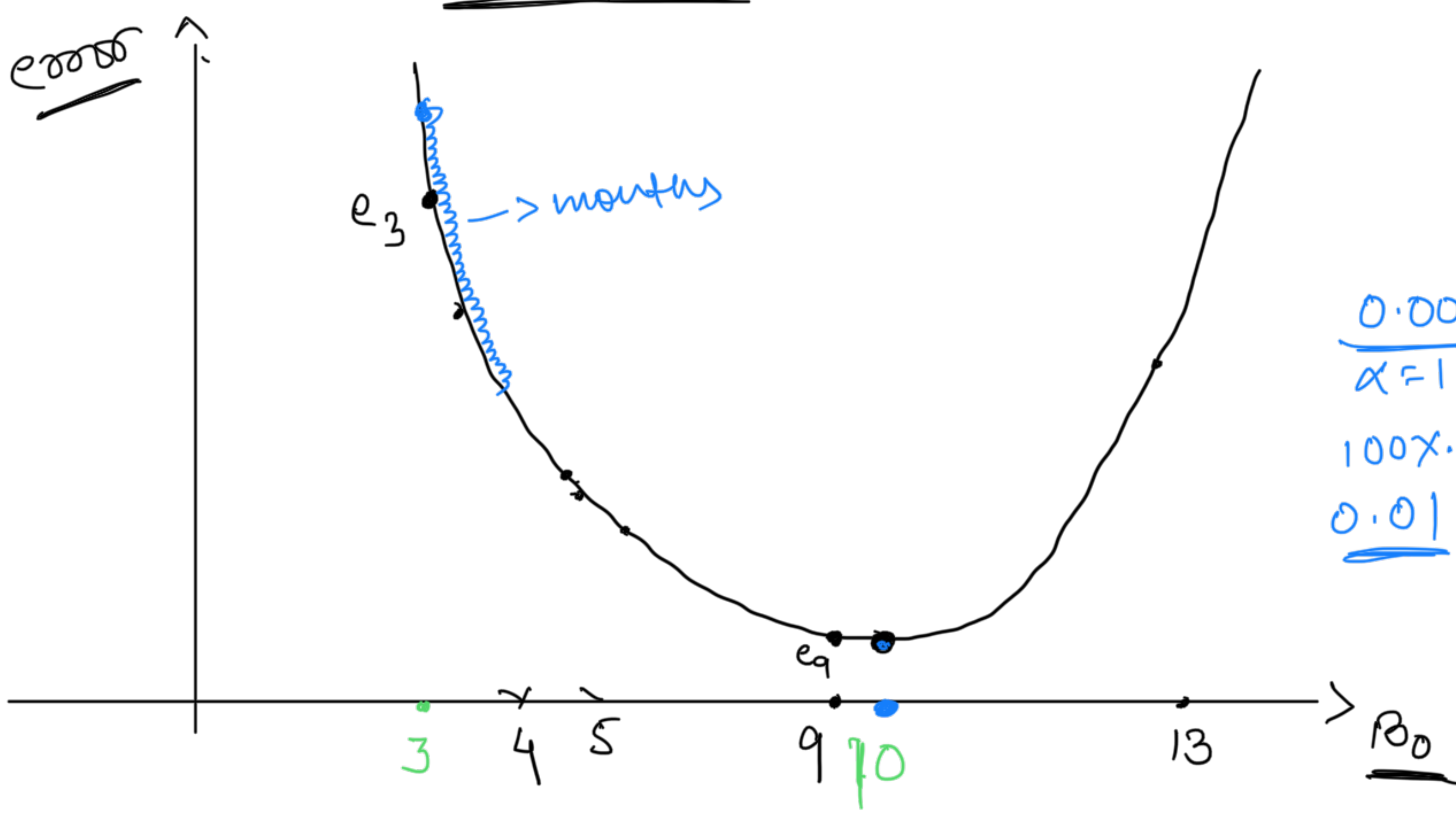
" $\sum_{i=1}^{n} (Y_i - \hat{B}_0 - \hat{B}_1 X_1)$

Solve this, B_0 and B_1 which is giving you the least error.

b. <u>Gradient Descent</u>

Obj:- Find $\underline{B_0}$ and $\underline{B_1}$ such that error is least lean.

β_0 vs Error



0.1

0.0001
$\alpha = 100$
$100 \times .0001$
0.01

$$\boxed{\text{New } \beta_0 = \text{Old } \beta_0 - \text{slope}}$$

-> sign

$$= 13 - (+K)$$

new β_0 $= 13 - K$

action -> move K steps towards left

$$\boxed{\beta_{0_{new}} = \beta_{0_{old}} - \alpha \cdot \frac{dE}{d\beta_0}} \quad -(1)$$

$$\boxed{\beta_{1_{new}} = \beta_{1_{old}} - \alpha \cdot \frac{dE}{d\beta_1}} \quad -(2)$$

1. Where are you starting??
2. How much iterations??
3. How much is learning rate??

Model evaluation :-

MAE
MSE
RMSE

Coefficient of determination :- (R^2)

Sales : 2 1 6 5 -3 7

Next value :- mean( ) ; my

Marketing spend : 10 12 6 2 3 5

Value :- better than my ✓
worst than my.

f1	Target
a	10
b	20
c	30
d	40
e	?? 25 (my)

For e ; using Mean pred
pred = 25

; use a ML model 1
pred = 30

Actual = 40

; use a ML model 2
pred = 32

When I am assuming the mean pred.

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

$$\text{Error} = \sum_{i=1} (y_i - \hat{y}_i)$$

$$TSS = \sum_{i=1}^{n} (y_i - m_y)^2$$

using a model preds $\gg m_y$

$$(y_i - \text{preds}) \ll (y_i - m_y)$$

$$RSS \ll TSS$$

Quantify how less is RSS to TSS. OR

how much the model is performing better than just mean prediction.

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

"explained"

$$SLR(X) \rightarrow Y$$

1. $TSS = RSS$

$$\frac{TSS - RSS}{TSS} = 0$$

2. $RSS = 0$

$$R^2 = 1$$

3. $0 < R^2 < 1$

$R^2 = 0.3$ $R^2 = 0.9$

30% of variance of o/p 90% of variance.

" " " " o/p