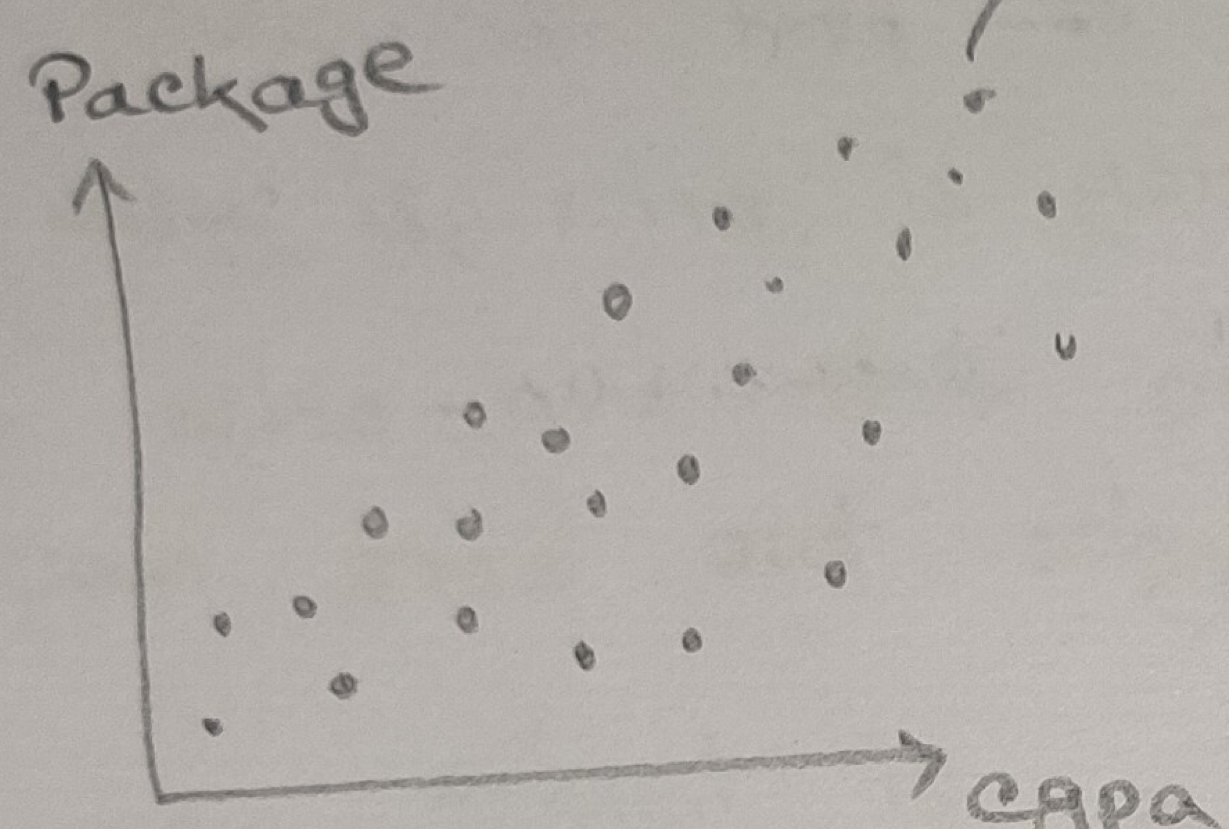
"Simple Linear Regression"

Landscape

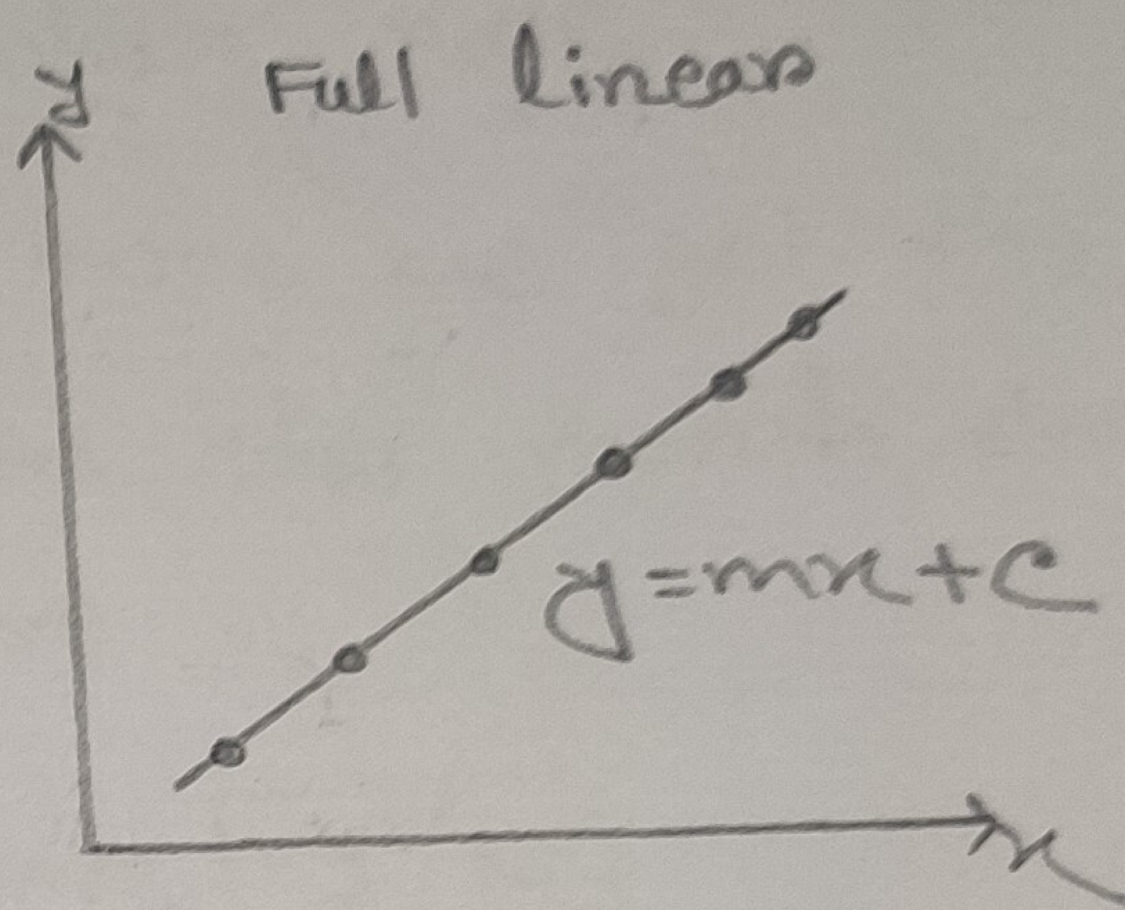
Types of Linear Reg.
1. simple
2. multivariable
3. polynomial
4. Regularization

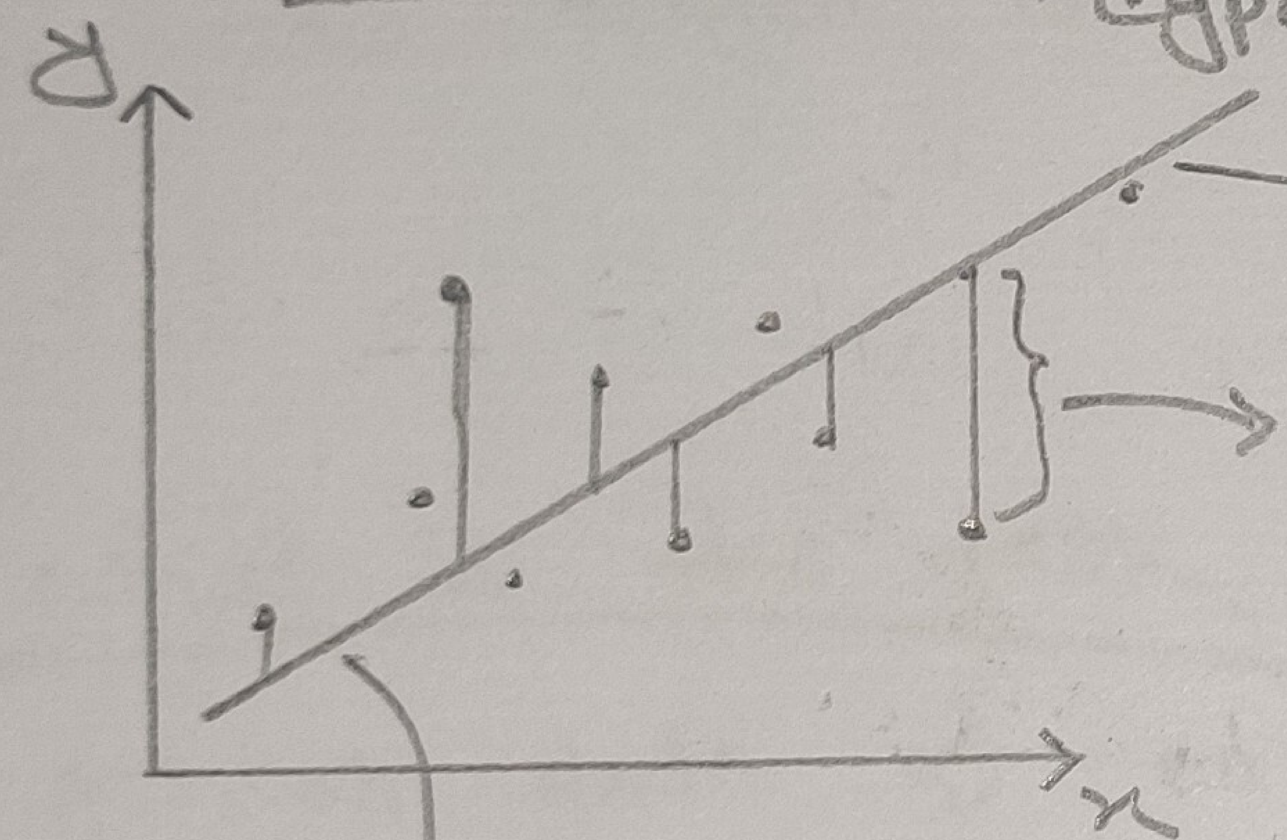
#simple LR

1 i/p → 1 o/p



→ sort of linear





→ best fit line

→ error value → linear regression tries to make a best fit line with minimum error

$\boxed{y = mx + b}$ → target is to find the value of m and b so that the error becomes minimum

Package = m * cgpa + b

b → offset

m ← weightage → how much package is depending on cgpa

$$\boxed{y = mx + b}$$

↓

we can do it in $\boxed{2 \text{ ways}}$

using some sort of formulas ← closed form solution

non-closed form solution

- It solves a given problem interms of functions (sin, log), variable constant and mathmatical operation
- using OLS (ordinary least square)
  ↓
  good when lower dimension

- approximation technique
- Gradient Desent method
  ↓
  easy to use in higher dimension

Sklearn → LinearRegression → use OLS

Sklearn → SGDRegression → use gradient decent

"Linear Regression using OLS"

$$\boxed{b = \bar{y} - m\bar{x}}$$

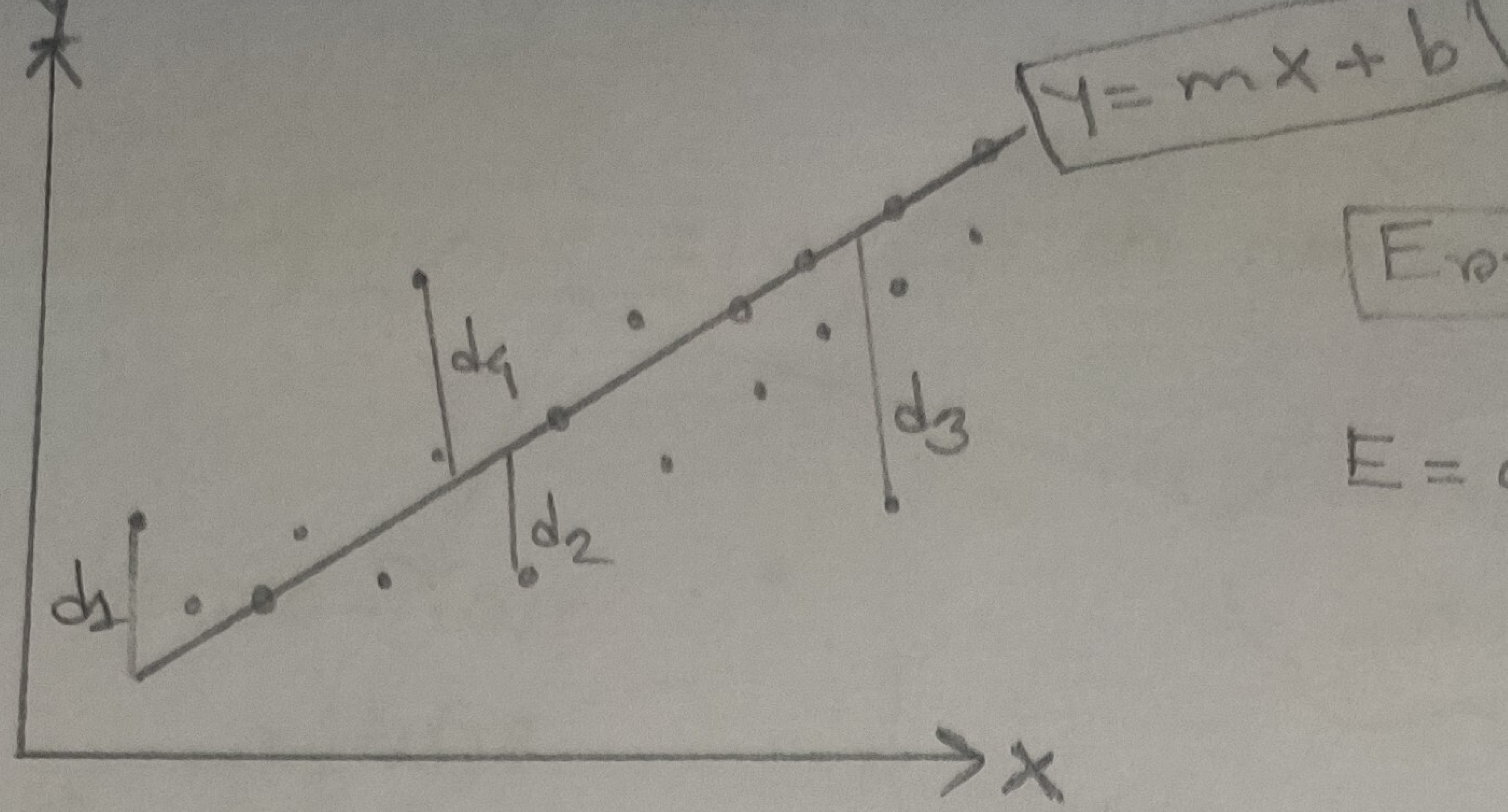
$$m = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$\bar{y}$ → mean of target column.

$\bar{x}$ → mean of input column

n → total point

Aristoderm



Error

$E = d_1 + d_2 + d_3 + d_4$ ✗

✗ $E = |d_1| + |d_2| + \dots$

(R1) → want to penalize outlier

(R2) → doing differentiation of mode function is not easy

$E = d_1^2 + d_2^2 + d_3^2 + \dots$

$$\boxed{E = \sum_{i=1}^{n} d_i^2} \rightarrow \text{loss function}$$

$d = y_i - \hat{y}_i$

actual

predicted

$$E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad \boxed{\hat{y}_i = mx_i + b}$$

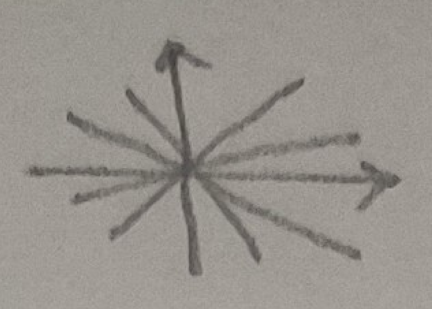
$$\boxed{E(m, b) = \sum_{i=1}^{n} (y_i - mx_i - b)^2} \rightarrow \text{now we need to minimize } E(m,b)$$

$E(m, b) \rightarrow m, b$ is the tuning parameter we can tune it. Both m, b or indivial m, b

what's impact on E?

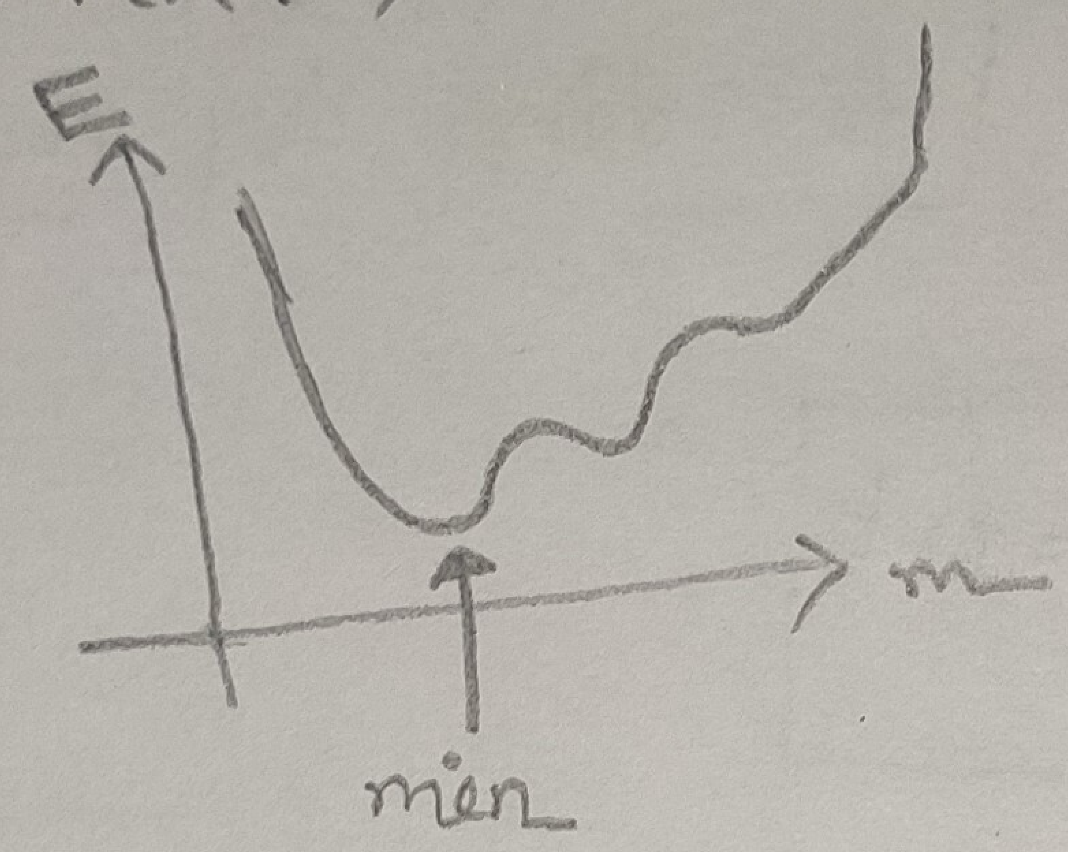
Impact of m only

let's $\boxed{b=0}$



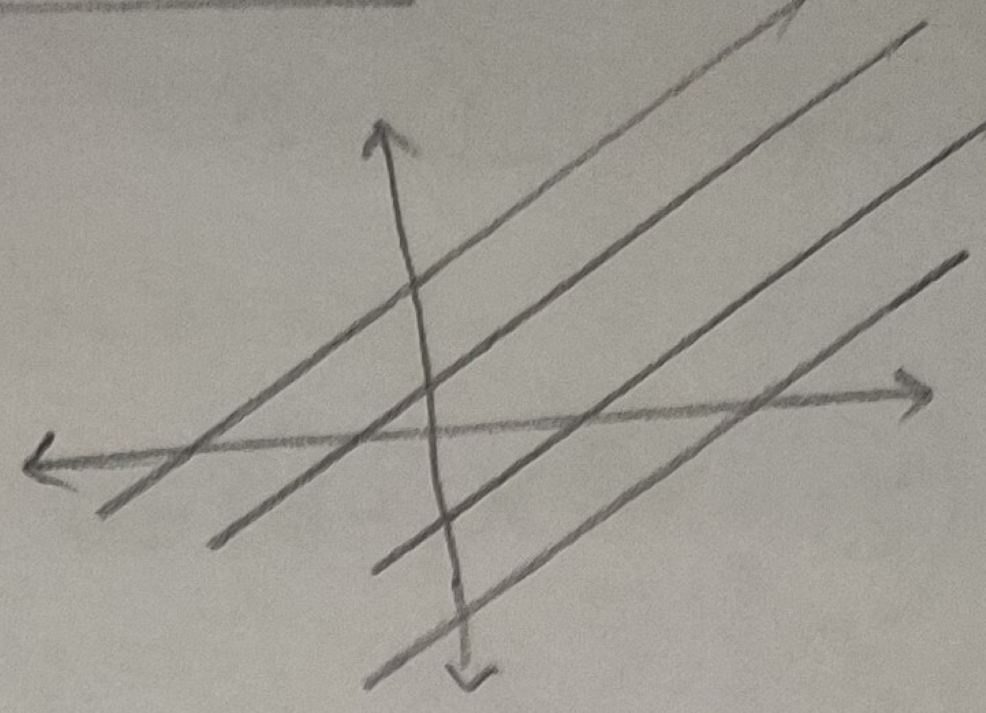
$$E(m)=\sum_{i=1}^{n}(\underbrace{y_i-mx_i}_{a})^2$$

$m\uparrow \Rightarrow mx\uparrow \Rightarrow a\downarrow \Rightarrow a^2\downarrow$

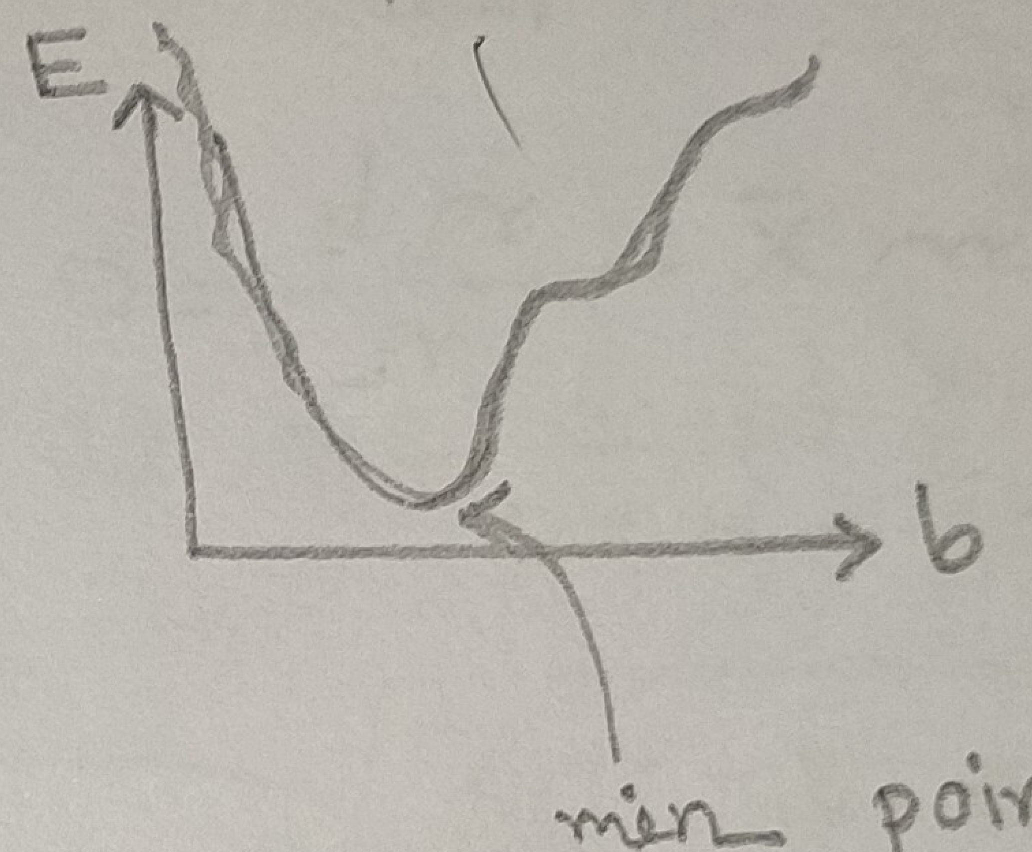


impact of b only

let m = 1 : $\tan\theta = 1 \rightarrow \theta = 45^\circ$



any line which make $\theta = 45^\circ$ with x axis



If we tune both m and b then we will get a 3d plot. there will be point where E(m,b) will be min

this point is call Minima → means 1st derivative will be 0

since we have m and b so both $\frac{\partial}{\partial m}$ and $\frac{\partial}{\partial b}$ will be 0

$$E(m,b)=\sum_{i=1}^{n}(y_i-mx_i-b)^2$$

$$\boxed{\frac{\partial E}{\partial m}=0} \qquad \boxed{\frac{\partial E}{\partial b}=0}$$

Aristoderm

$$\frac{\partial E}{\partial b} = 0$$

$$\sum_{i=1}^{n} \frac{\partial}{\partial m} (y_i - m x_i - b)^2 = 0$$

$$\Rightarrow 2 \sum_{i=1}^{n} (y_i - m x_i - b) = 0$$

$$\Rightarrow \sum_{i=1}^{n} \frac{y_i}{n} - m \sum_{i=1}^{n} \frac{x_i}{n} - \sum_{i=1}^{n} \frac{b}{n} = 0$$ [dividing by n]

$$\Rightarrow \bar{y} - m\bar{x} - \frac{nb}{n} = 0 \qquad \boxed{\sum_{i=1}^{n} b = nb}$$

$$\Rightarrow \bar{y} - m\bar{x} - b = 0$$

$$\Rightarrow \boxed{b = \bar{y} - m\bar{x}} \longrightarrow E(m,b) = \sum_{i=1}^{n} (y_i - m x_i - \bar{y} + m\bar{x})^2$$

$$\frac{\partial E}{\partial m} = 0$$

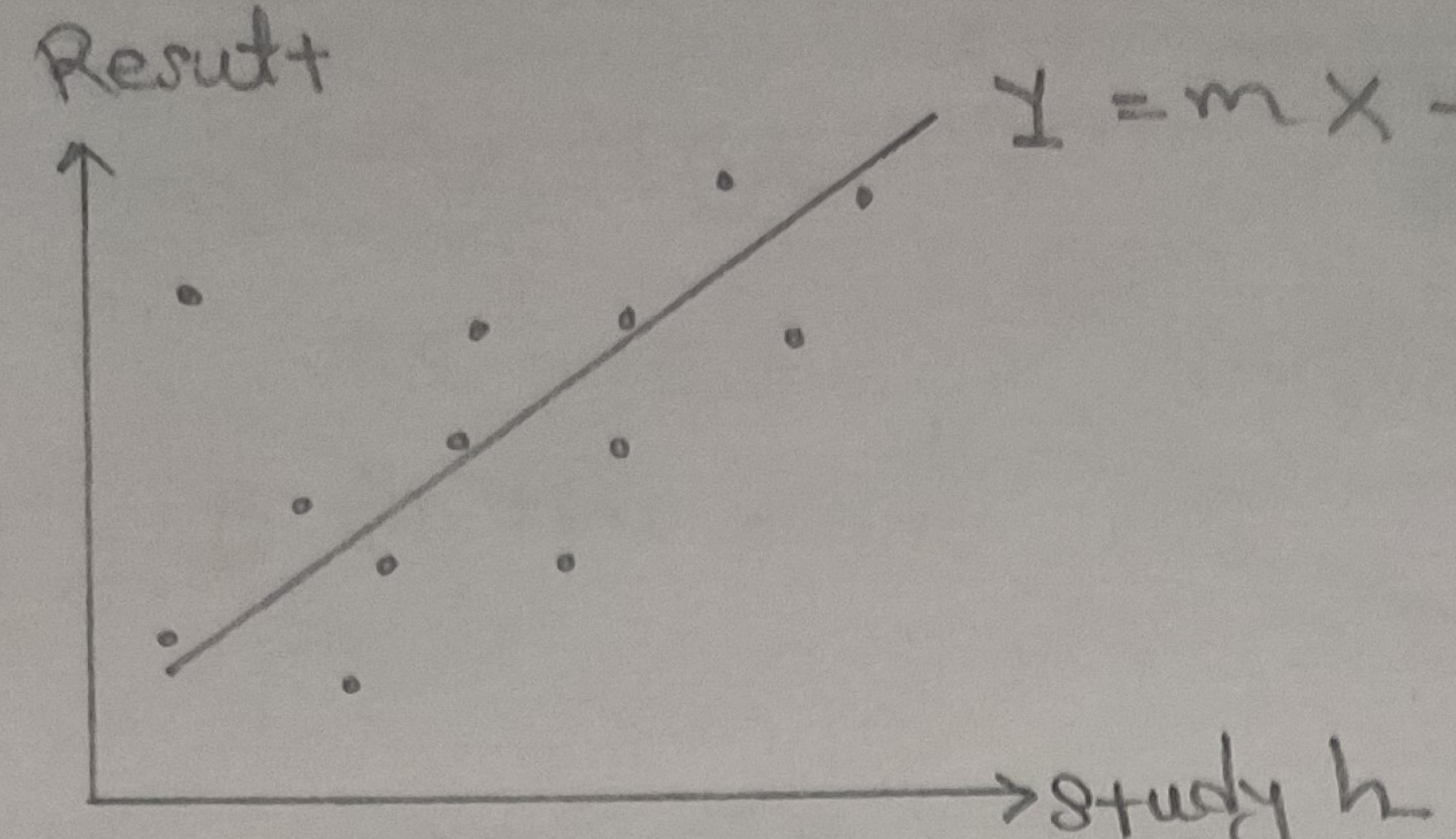
$$\Rightarrow 2 \sum_{i=1}^{n} (y_i - m x_i - \bar{y} + m\bar{x})(-x_i + \bar{x}) = 0$$

$$\Rightarrow \left\{ \sum_{i=1}^{n} (y_i - \bar{y}) - m \sum_{i=1}^{n} (x_i - \bar{x}) \right\} (x_i - \bar{x}) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) = m \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\Rightarrow m = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

"Simple Linear Regression using Gradient Descent"



cost function → Error Function (MSE)

$$E(m,b) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

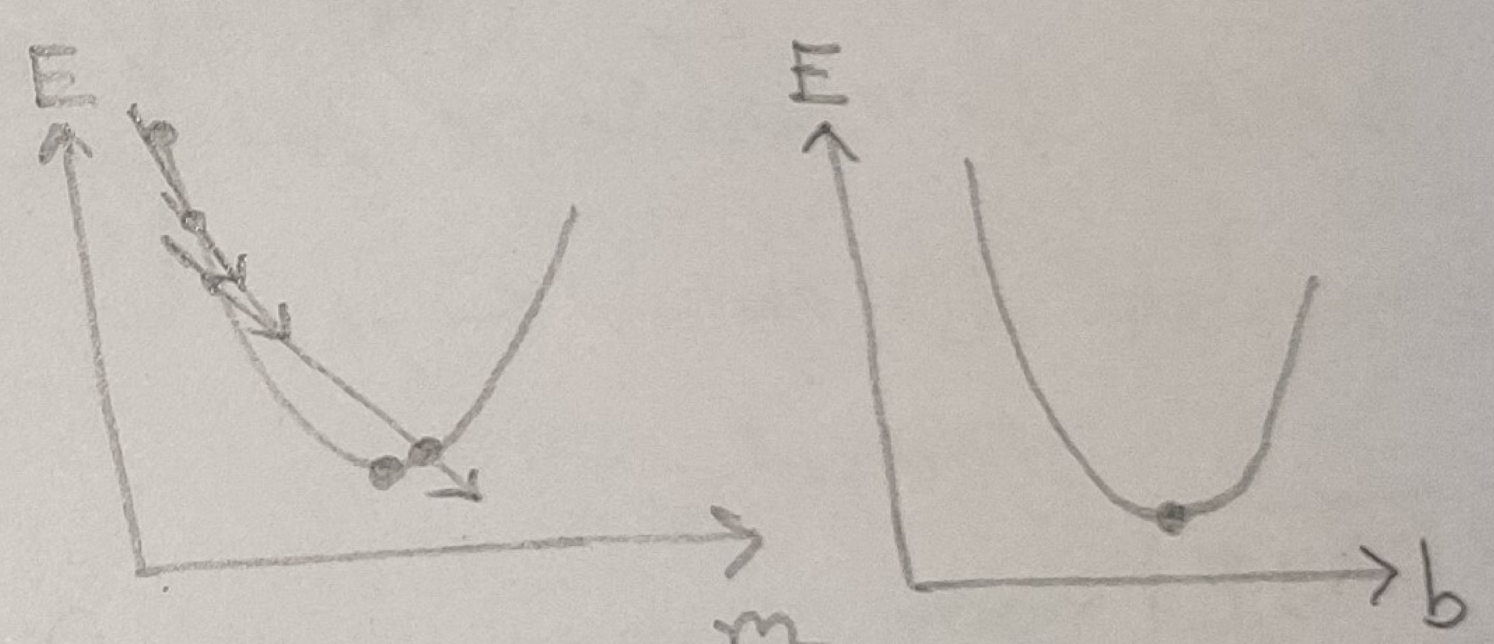
$$E(m,b) = \frac{1}{n}\sum\{y_i - mx_i - b\}^2$$

Mean squre Error → need to minimize this

$$\frac{\partial E}{\partial m} = \frac{1}{n}\sum_{i=1}^{n} 2(y_i - mx_i - b)(-x_i)$$

our task is to minimize th E. If we could find the value and direction of the gradient

Partial derivative of E will give direction of steepest ascend with respect to m and b so how can we change m v to maximally increase e. Then will go the opposite and b. to minimize.



$$\frac{\partial E}{\partial m} = \frac{-2}{n}\sum_{i=1}^{n} x_i(y_i - mx_i - b)$$

$$\frac{\partial E}{\partial b} = \frac{-2}{n}\sum_{i=1}^{n}(y_i - mx_i - b)$$

$$m = m - LR * \frac{\partial E}{\partial m}$$

$$b = b - LR * \frac{\partial E}{\partial b}$$

$$\frac{\partial E}{\partial b} = \frac{-2}{n}\sum_{i=1}^{n}(y_i - (mx_i + b))$$

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

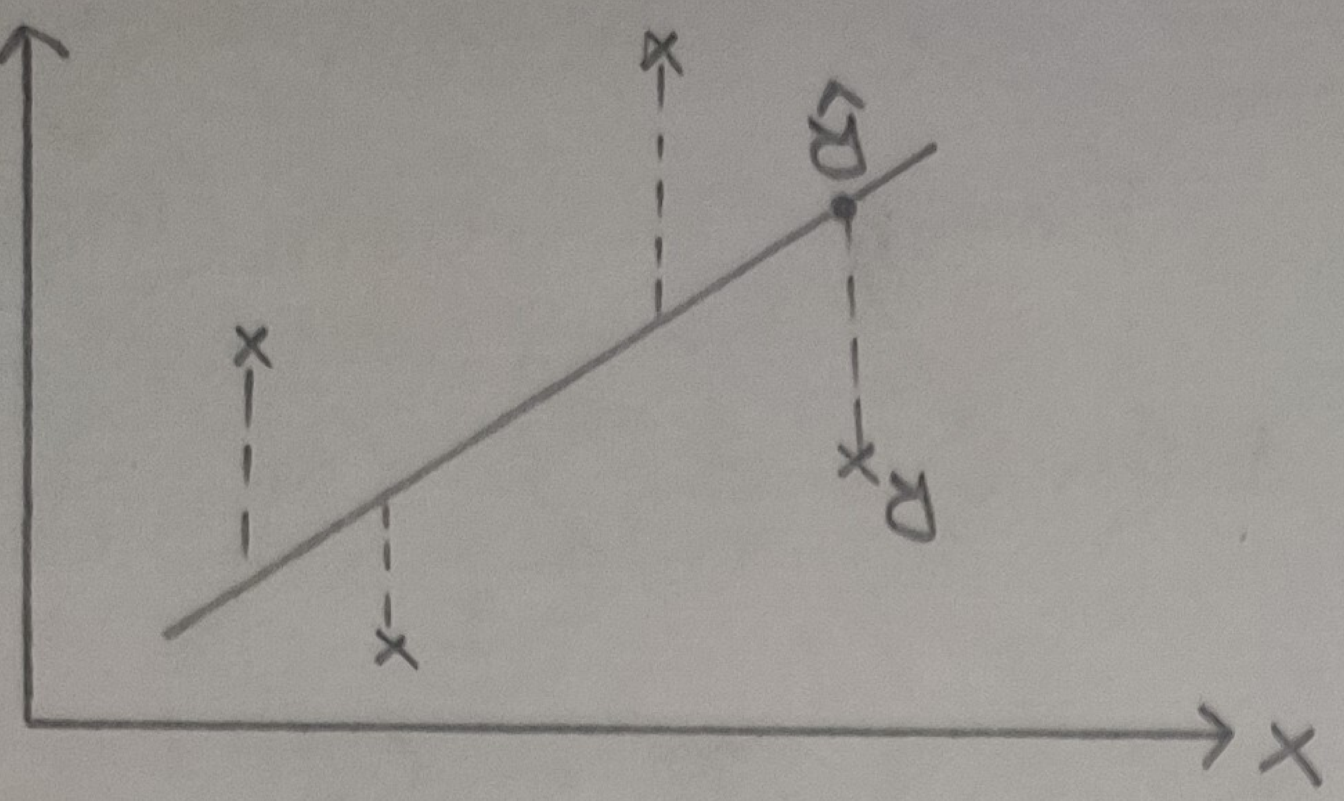
dot product : $\sum_{i=1}^{n} x_i \cdot \text{error}$

$= np.dot(x_i, \text{error})$

$$\frac{\partial E}{\partial m} = \frac{-2}{n}\sum_{i=1}^{n} x_i \underbrace{(y_i - \hat{y}_i)}_{\text{error}}$$

Aristoderm

MAE Mean Absoulte Error



$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}$$

Loss

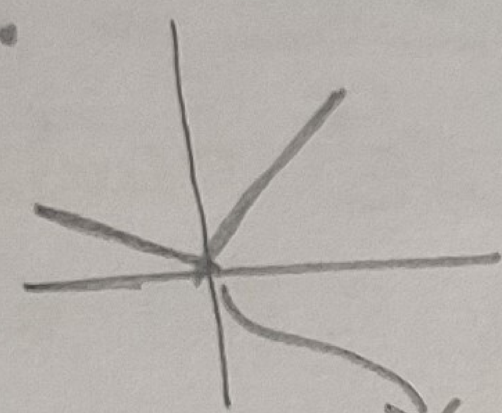
unit = unit of y

Advance:

① Same unit

② Robust to outlier (not that much effect)

Disadvantage:



at zero graph is not differentiable

MSE mean squared error

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

advantage

① can be used as loss function cause it's differentiable

Disadvantage

① unit = $(\text{target unit})^2$

② hugely effected by outlier

RMSE root mean square error

$$rms = \sqrt{mSE}$$

$\boxed{\text{R2 Score}}$ MAE, MSE, RMSE → context depended

context independent. → goodness of fit

$$R2 = 1 - \frac{SS_R}{SS_m}$$

SS_R → sum of squared error in Regression line

SS_m → sum of squared error in the mean line

$R2 = 0 \rightarrow SS_R = SS_M \rightarrow$ worst

$R2 = 1 \rightarrow SS_R <<< SS_m \rightarrow$ Best

$R2 < 0 \rightarrow SS_m >> SS_R \rightarrow$ worst worst

cg	LPA

→ $R2 = 0.80$ → means cg explain 80% of varience of lpa

Sometimes by adding irrelevent feature R2 score gets better which should not. That's where we use adjusted R2 score.

$$R^2_{adj} = 1 - \left[\frac{(1-R2)(n-1)}{(n-1-k)}\right]$$

R_2 → R2 score

n → number of Rows

k → num of independent columns.

Aristoderm