Simple Linear Regression

- easy to understand.
- Foundation Algorithm

LR → Supervised Machine learning Algorithm

Regression

Simple

- 1 input Column
- 1 output Column

cgpa | package

Multiple LR

- Multiple input Column

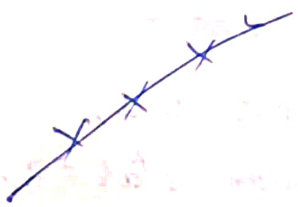
cgpa | gender | 12th | state | package

Polynomial LR

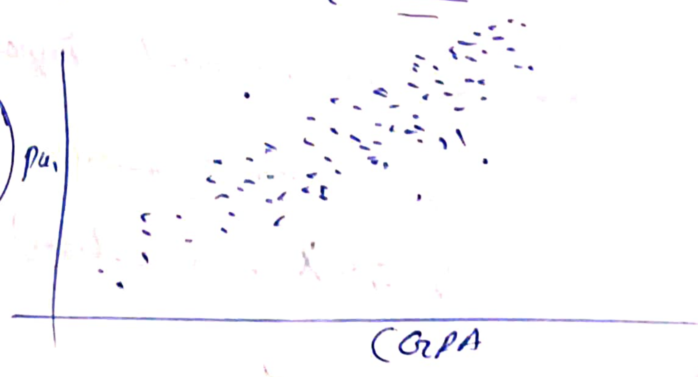
cgpa	package
4.1	3.5
4.7	1.2
8.4	4.2
8.1	3.4

cgpa → model → package

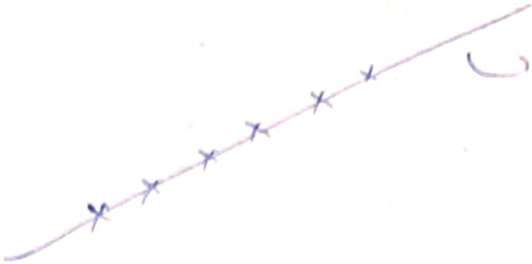
first, plot the Data



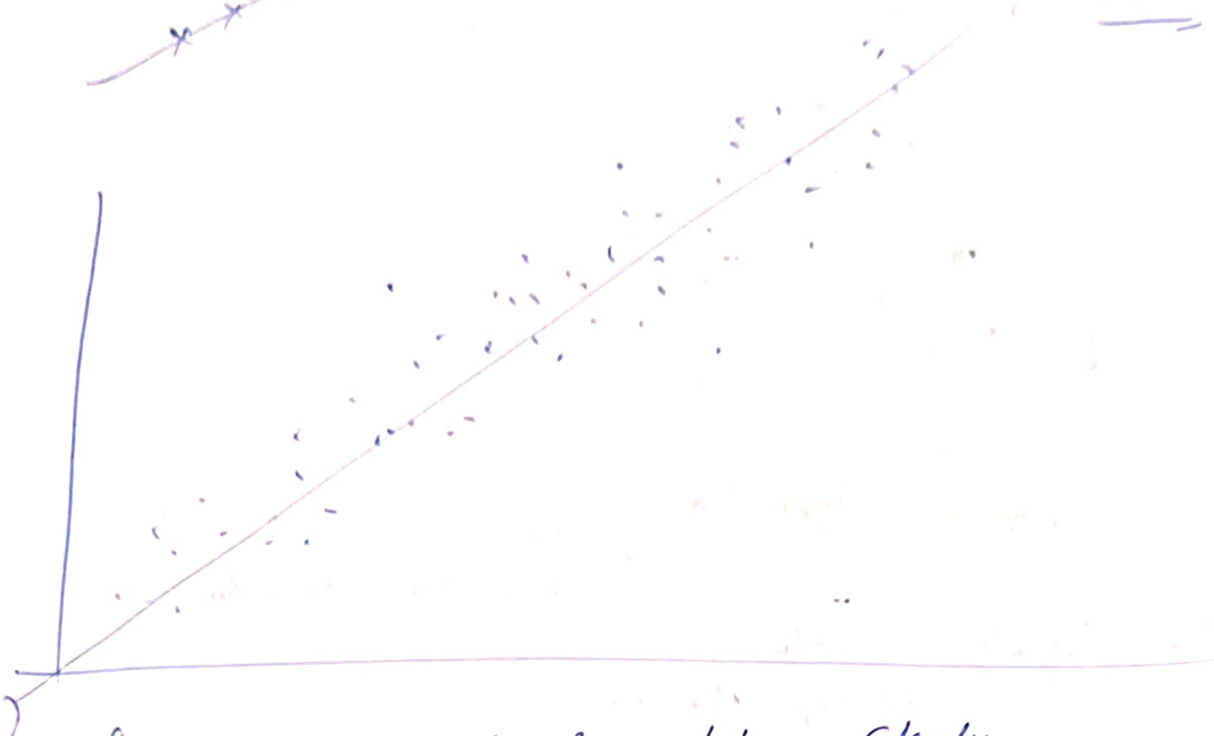
Real world Dataset → Sort of Linear Data



stochastic error → jo determin nahi kar sakte hai



↳ perfect linear Data
$y = mx + c$
↳ Very easy to Determine



best fit line ↳ Saare points se ekdam Clearly. Pass kare.

↳ in points pe minimum Galti kare
↳ find that (m, b).

$m \rightarrow$ slope
$b \rightarrow$ y-intercept

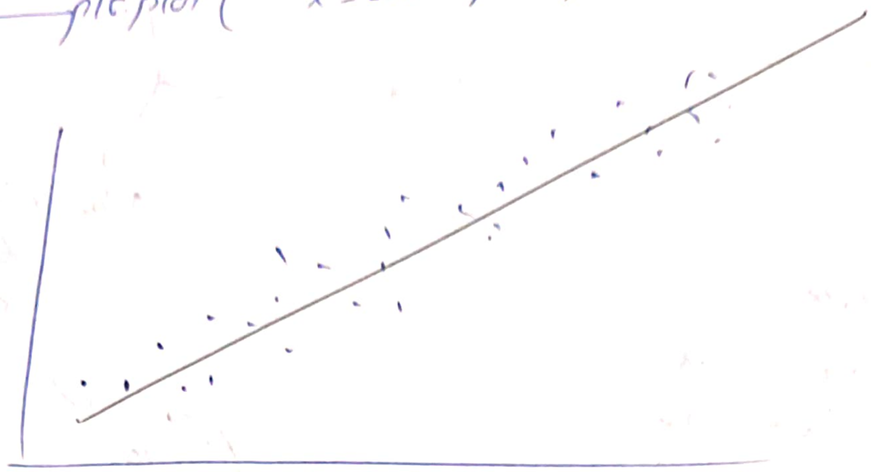
```
X = df.iloc[ :, 0:1]
y = df.iloc[ :, -1]

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split
                       ( X, y, test_size = 0.2,
                              random_state = 2)

from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(X_train, y_train)
```

Now ~~lr.predict (X_test.iloc[0])~~

lr.predict (X_test.iloc[0].values.reshape(1,1))

→ See the regression line

plot { —— plt.plot(X_test, lr.predict(X_test), color = 'red'))

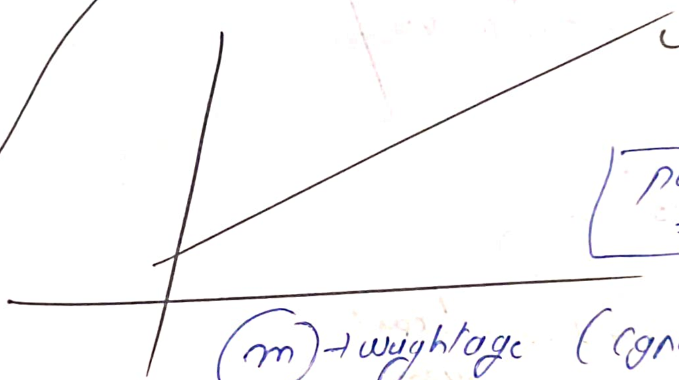


m = lr.coef_ → Slope

b = lr.intercept_

$y = mx + b$

Intuition



$y = mx + b$

package = m × cgpa + b

(m) → weightage (cgpa pe kitna depend)

m↓ (↑) weightage of cgpa on package

exp 0 | pac

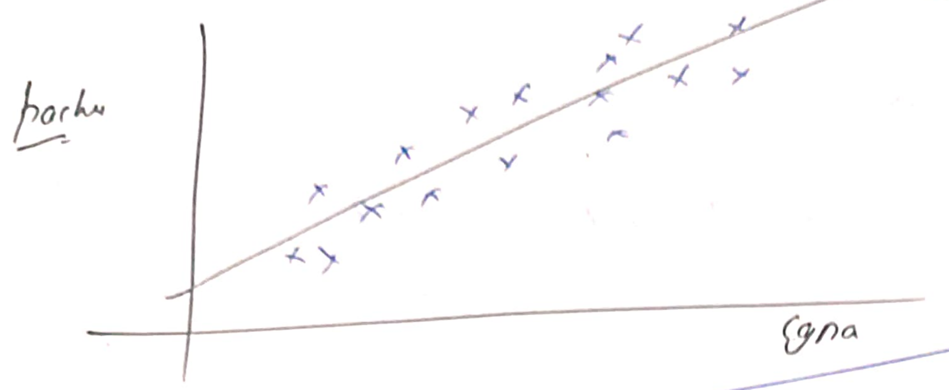
b = 0

package = m × cgpa

offset ki agar m = 0 still kuch Logo

SIMPLE LINEAR REGRESSION

MATHEMATICAL FORMULATION



↳ best fit line
↳ $y = mx + b$

(m, b)

Closed form Solution

(ek formula bana sakte ho bina Differentiation or Integration ke)

then it is a Closed form Solution.

like
$$\boxed{x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}}$$

Direct formula
↳ OLS
Oridinary Least Square

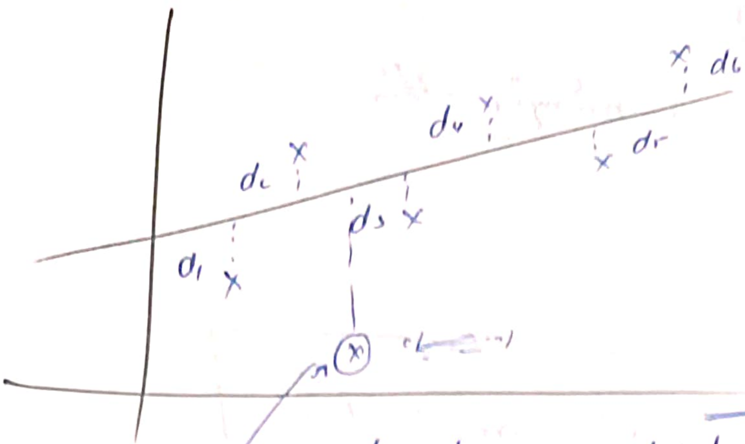
Non-closed form Solution

(Gradient Descent Technique
↳ Higher Dimension main jyada chalta Hai.

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

$\bar{x} \rightarrow$ mean
$\bar{y} \rightarrow$

Internally Sklearn
LR → OLS impl
SGD Regr → Gradient Descent implemented



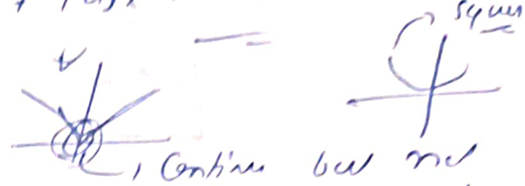
(m, b) ki value nikalna

$E = d_1 + d_2 + d_3 + \dots + d_n$) square karda

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

Why not mod and square $E = |d_1| + |d_2| + |d_3| + \dots$

↳ Reason - 1



↳ Continue but not diffr at origin

↳ Reason - 2

{ Outlier ko penalise karna chahte he

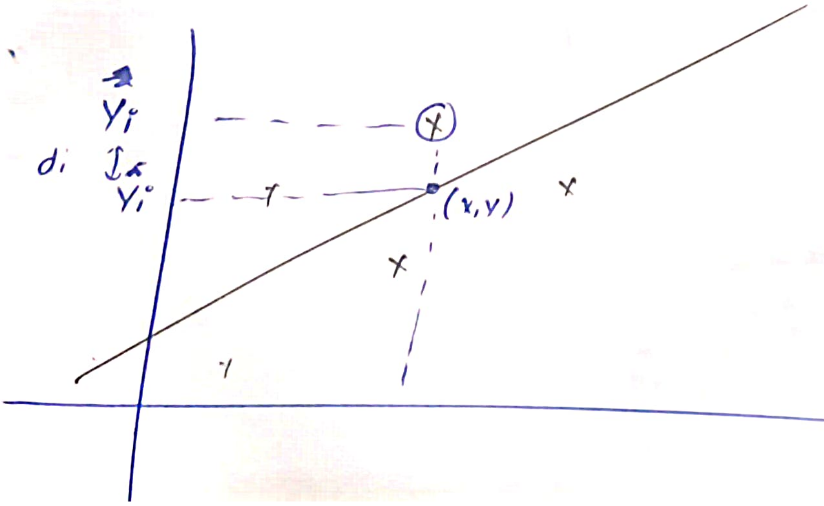
↳ refer to reducing the influence/weight of outliers when training a model so they don't spoil the Results.

↳ jo dur ka error hai use exaggerate karna

Hum Squared error use ↳ penalises outlier more strongly
↳ is differentiable
↳ gives a clean Mathematical Solution

Now $\boxed{E = \sum_{i=1}^{n} d_i^2}$ → error function represented by (J) also

↑ minimize karna aisa (m, b) chahiye.



$\boxed{d_i = y_i - \hat{y}_i}$

$Y = mx + b$

$$\boxed{E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$ → minimize this

$$\boxed{\hat{y}_i = m x_i + b}$$

$$* \boxed{E_{(m,b)} = \sum_{i=1}^{n} (y_i - m x_i - b)^2}$$

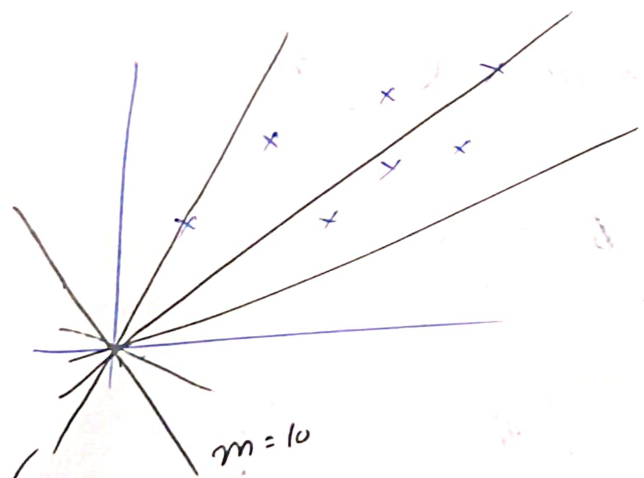
$$\boxed{E(m,b) = \sum_{i=1}^{n} (y_i - m x_i - b)^2}$$ minimize

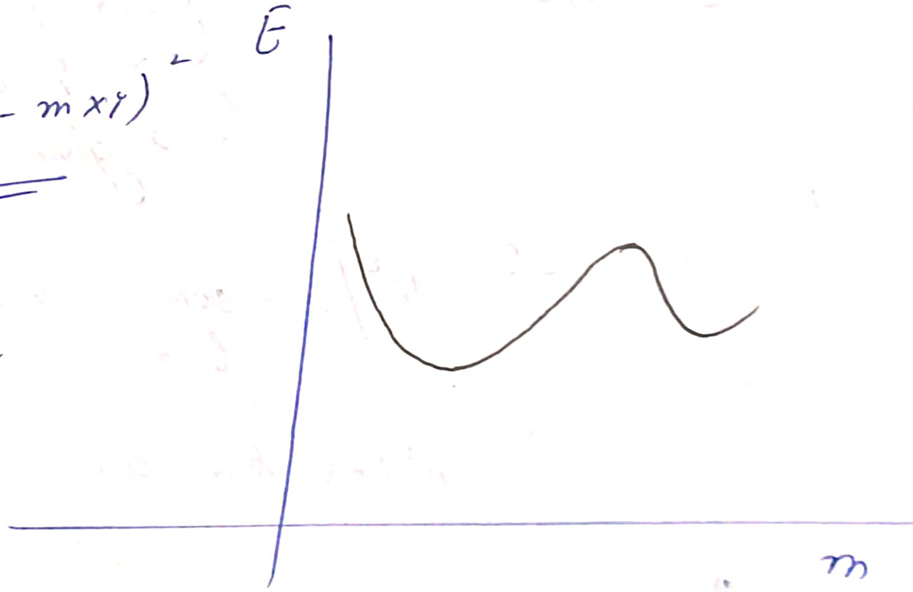
$y = f(x)$ x in setting mean for y.

(m, b) + an setting knob.

m $b = 0$

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





Only m change

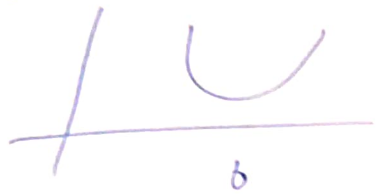
∴ rotate

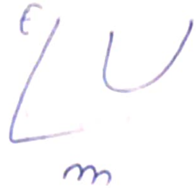
m ko ghumane se

$m = 1$ b

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

↳ Only Shifting line

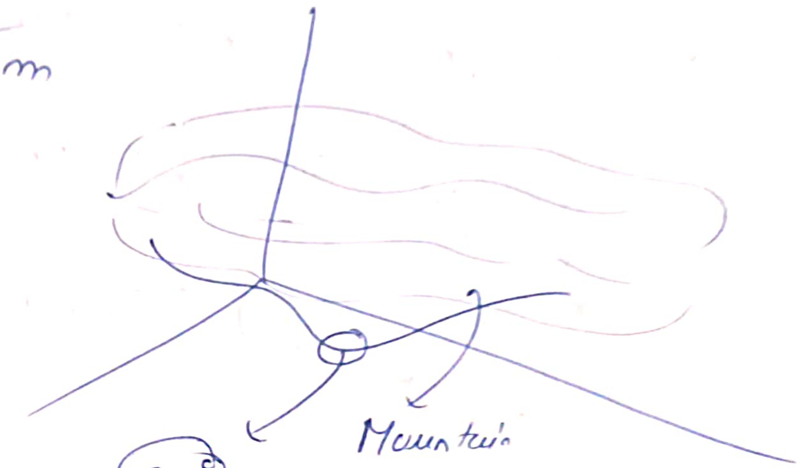




$E(x)$

↳ $f(x, y)$

∴ partial Derivative karna padega.



slope = 0 — min

derivative to zero — maxima / minima

Mountain ka step

$$\frac{\partial E}{\partial b} = \frac{\partial}{\partial b}\left(\sum \quad\right) = 0$$

$$= \sum \frac{\partial}{\partial b}(y_i - m x_i - b)^2$$

$$= \sum 2(y_i - m x_i - b)(-1) = 0$$

So $\sum (y_i - m x_i - b) = 0$

$$\frac{\sum y_i}{n} - \frac{\sum m x_i}{n} - \frac{\sum b}{n} = 0$$

$$\frac{b + b + b \cdots n \text{ times}}{n}$$

$$\frac{nb}{n} = b$$

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

b ko fit karenge Seead wale m ko nikal Denge.

$$E = \sum (y_i - mx_i - \bar{y} + m\bar{x})^2$$

$$\frac{\partial E}{\partial m} = \sum \frac{\partial}{\partial m} (y_i - mx_i - \bar{y} + m\bar{x})^2 = 0$$

$$= \sum 2(y_i - mx_i - \bar{y} + m\bar{x})((-x_i) + \cancel{m}\,\bar{x}) = 0$$

$$\Rightarrow \sum -2(y_i - mx_i - \bar{y} + m\bar{x})(+x_i - \bar{x}) = 0$$

$$\Rightarrow \sum ((y_i - \bar{y}) - m(x_i - \bar{x})] (x_i - \bar{x})$$

$$\Rightarrow \sum ((y_i - \bar{y})(x_i - \bar{x}) - m(x_i - \bar{x})^2] = 0$$

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