LR — SL

SL → Regr. | Classification

Regr. → O/p → numerical.

LR
→ 1) Simple
→ multiple
→ polynomial
→ Regularization

→ Data is not linear.
Different regression model.

Simple → one row and one column

Simple regression

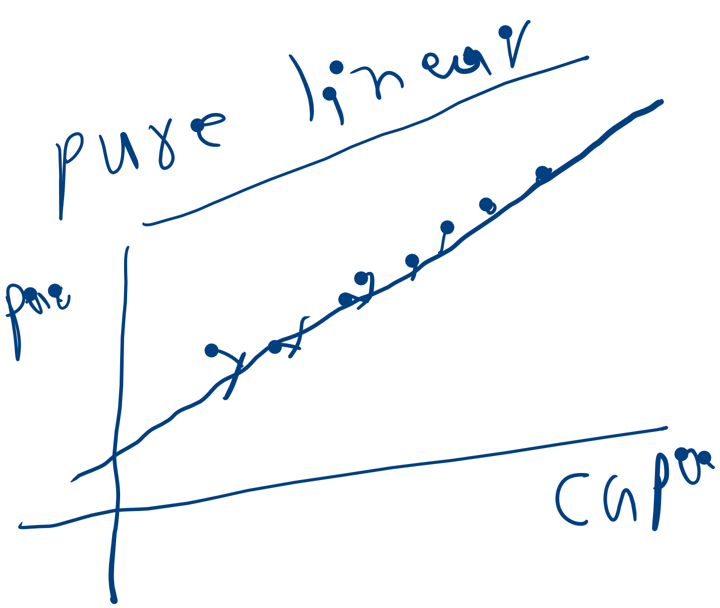
cgpa	package
7.1	3.5
4.7	1.2
8.9	4.2
8.1	3.9

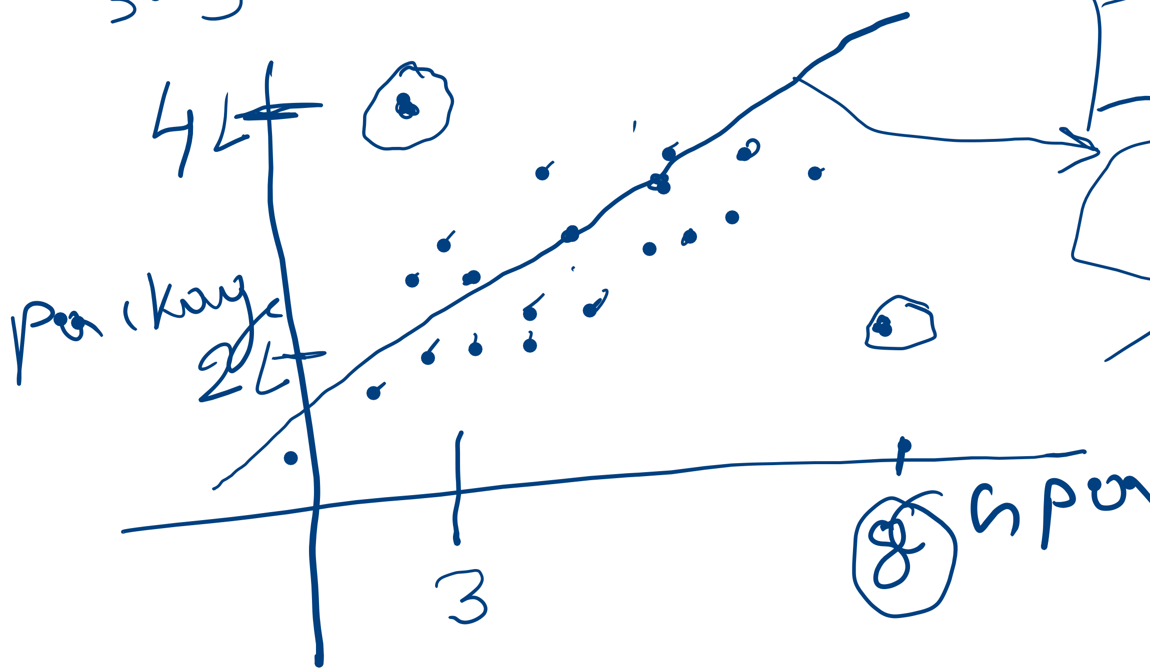
cgpa → model → package

Application

website

Sort of Linear





Stochastic error

Certain randomness output

Calculate m and b?

m, b

1 closed-form solution

2 non-closed solution

+, −, ×, ÷, fn log, exp, ...

× limit, diff.. infer

approximation based integration & differentiation

Gradient descent.

Closed-formed solution

- Direct formula
  - m, b

OLS → ordinary least square

Sklearn → OLS

SGDRegression — GD

Direct formula

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

$\bar{y}$ mean

$\bar{x}$

x → cgpa

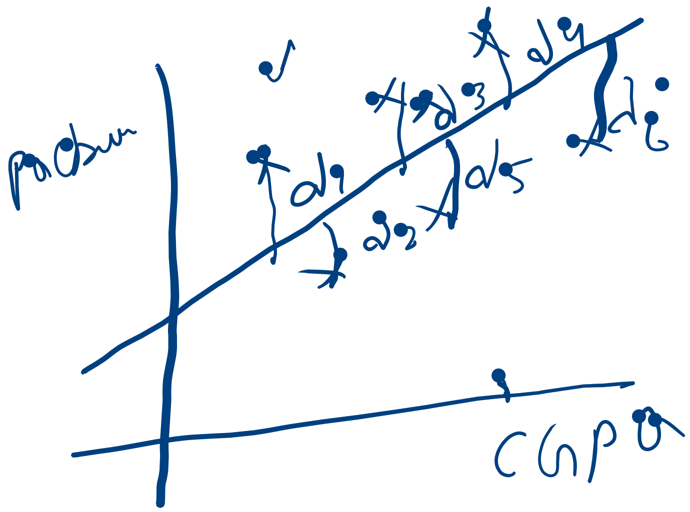
y → package

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

$x = \text{GPA}$ $\quad$ $\bar{x}$ } mean

$y = \text{package}$ $\quad$ $\bar{y}$

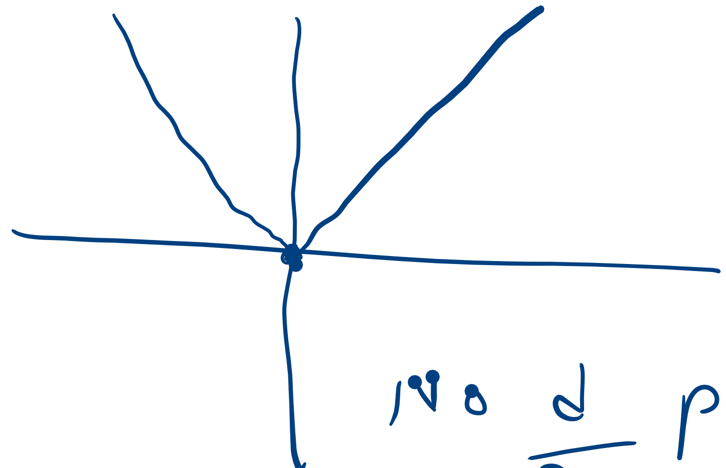
$$m = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$



$$E = d_1 + d_2 + d_3 + \dots d_n$$

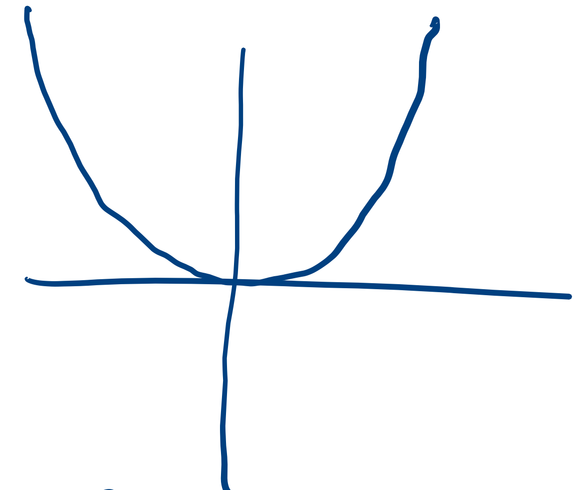
$$E = d_1^2 + d_2^2 + d_3^2 + d_n^2$$

mod



$$\frac{dx^2}{\partial x} =$$

No $\frac{d}{\partial x}$ possible



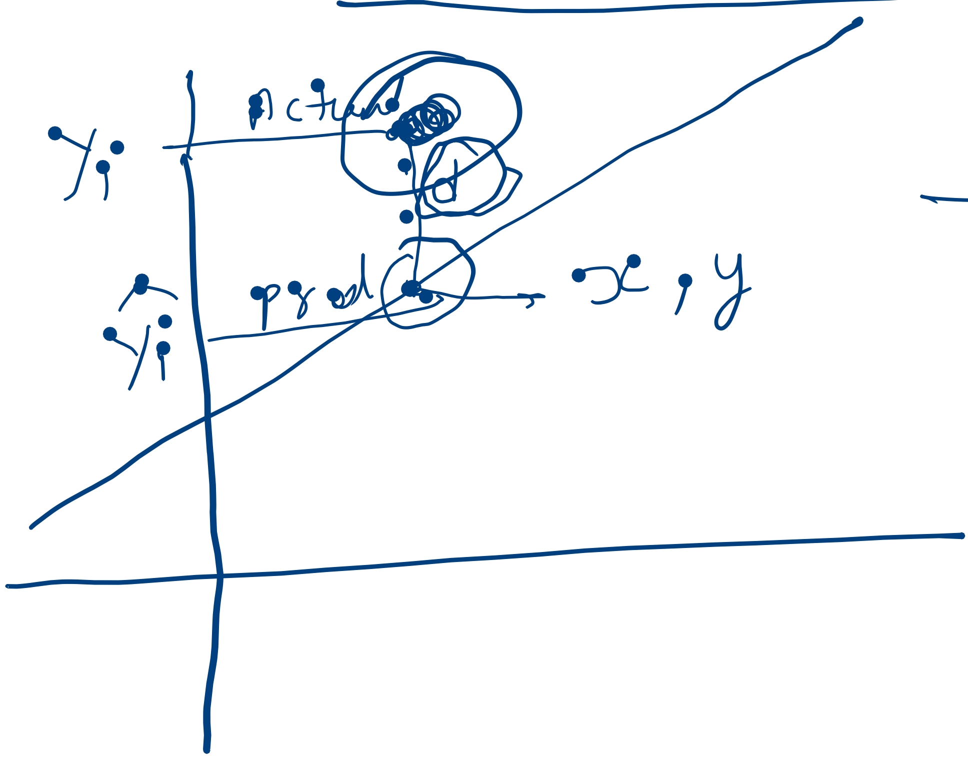
Squ

$$E = d_1^2 + d_2^2 + d_3^2 - \cdots d_n^2$$

$$\boxed{E = \sum_{i=1}^{n} d_i^2}$$

y_i = actual

$\hat{y}_i$ = predicted



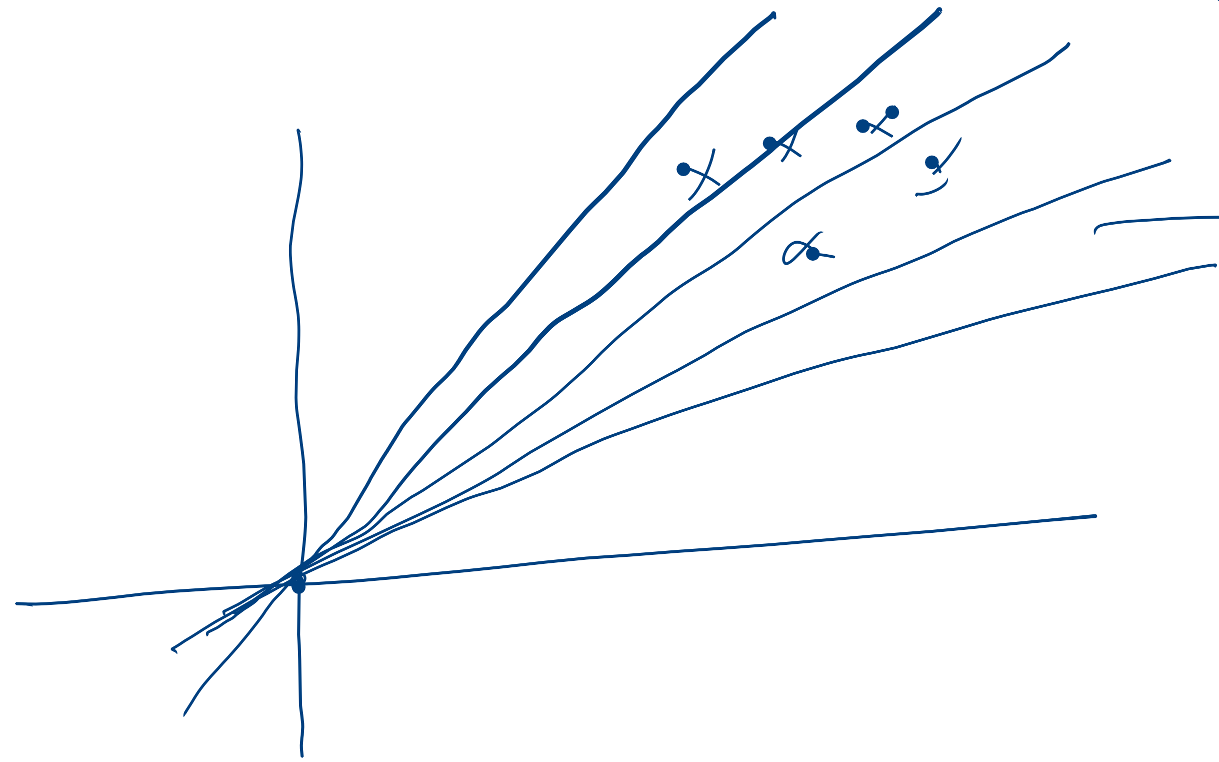
$$E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

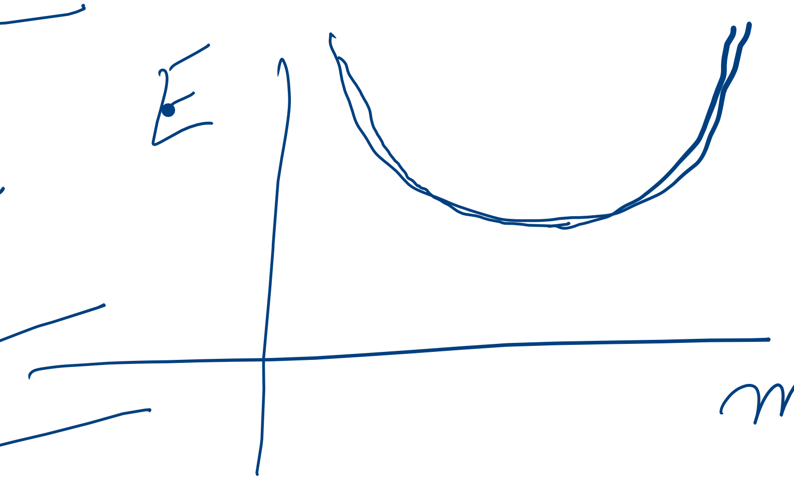
m, b hidd

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

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

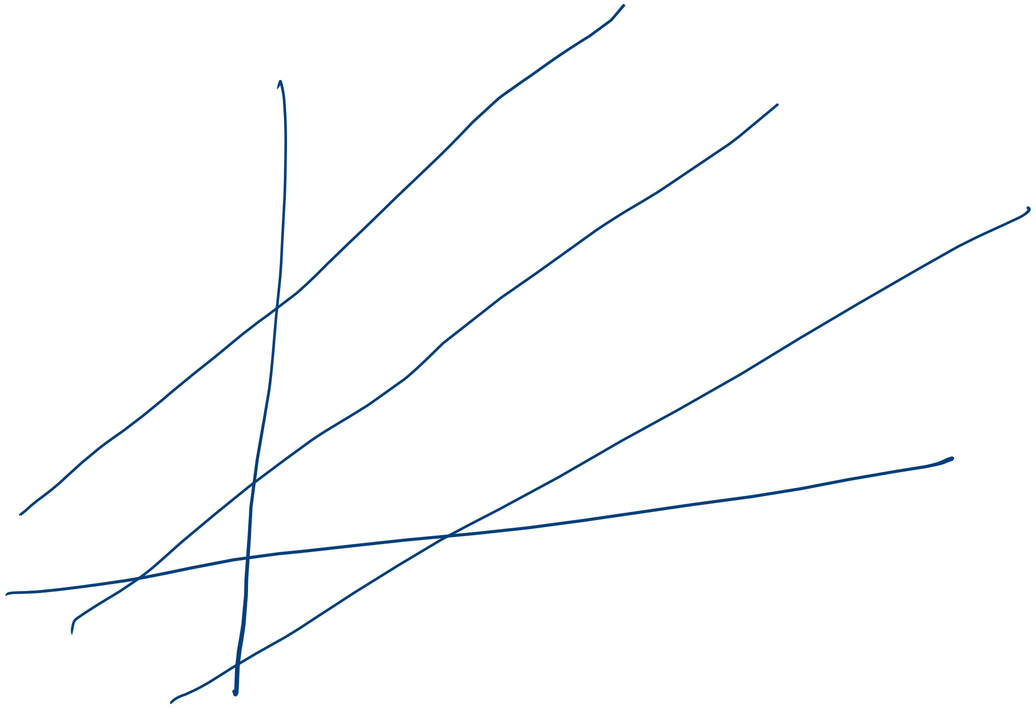
$b = 0$

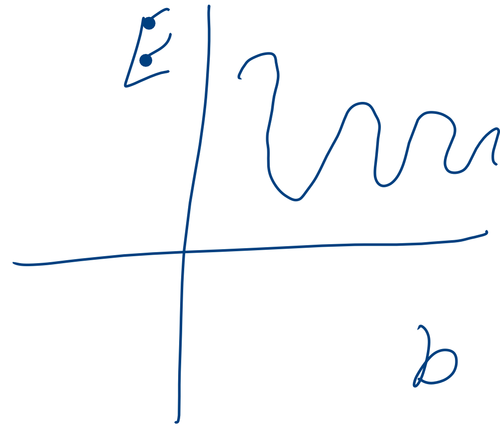




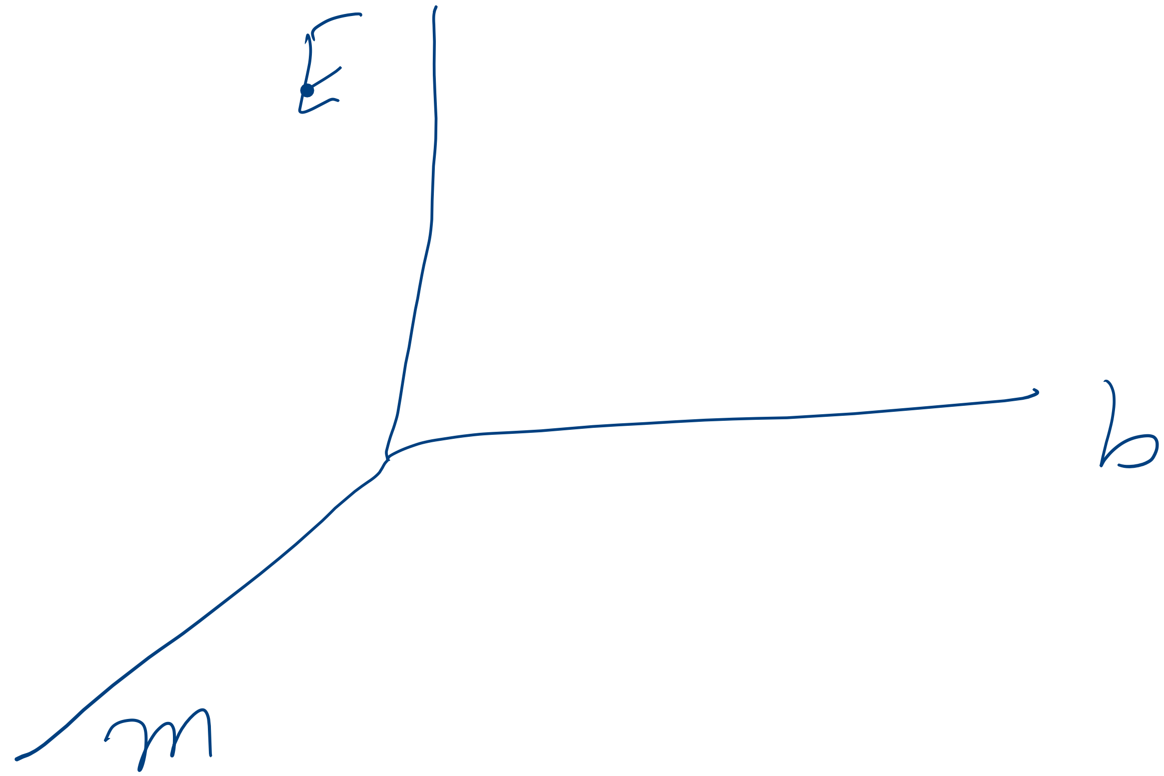
m

m = 1





m, b



$$\boxed{\frac{\partial E}{\partial m} = 0} \qquad \boxed{\frac{\partial E}{\partial b} = 0} \qquad \frac{\partial x^2}{\partial x} = 2x$$

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

chain rule

$$= \sum 2(y_i - mx_i - b) \cdot (-1)$$

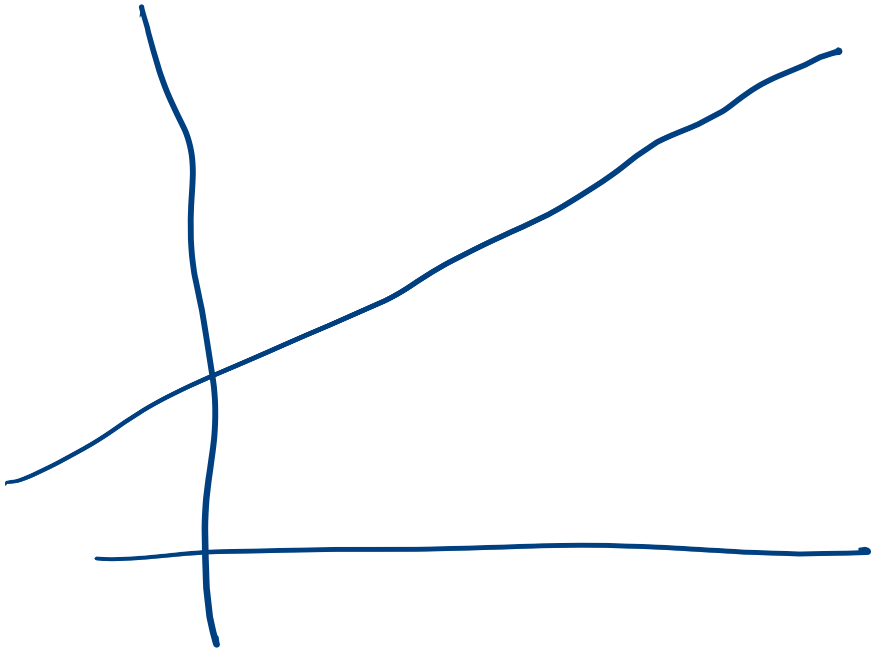
$$0 = \sum -2(y_i - mx_i - b)$$

$$\frac{0}{-2} \qquad -2$$

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

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

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



SLR

$\boxed{m, b}$ → offset

slope

$\boxed{\text{model}}$ → How → Regression

metrics

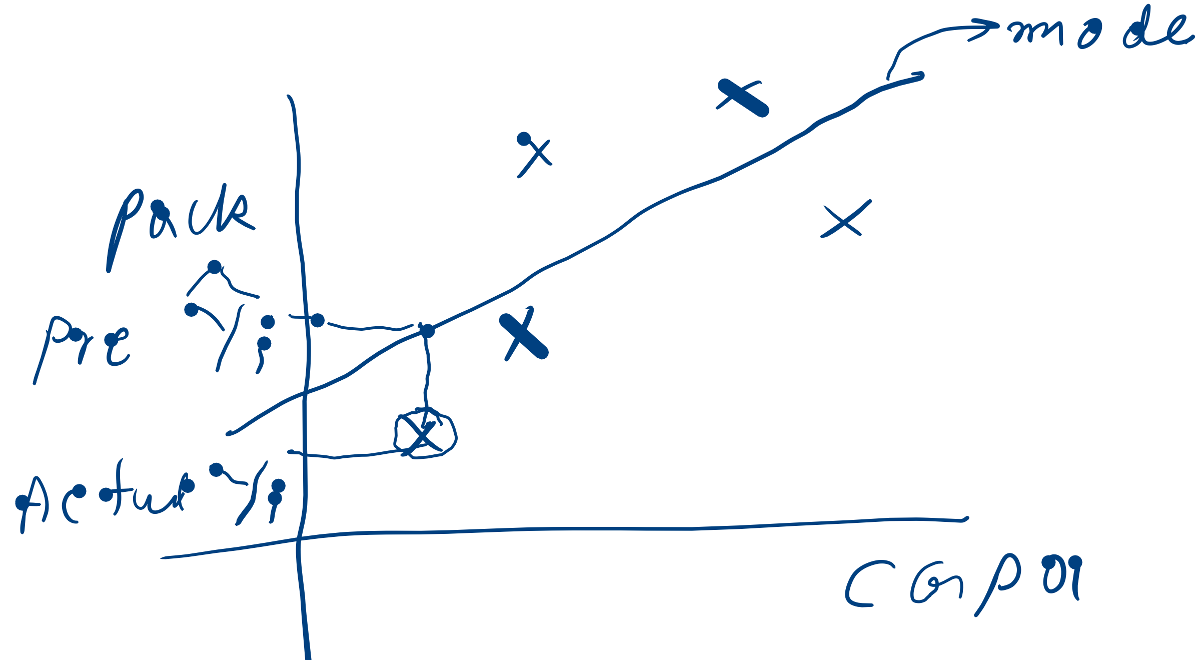
Efficien

Regression metrics

- MAE → mean absolute error
- MSE — mean square error
- RMSE → Root MSE
- R2 Score — Coefficient of determination / goodness of fit
- Adjusted R2 score

m

- **MAE** → mean Absolute error



$$\frac{|y_1-\hat{y}_1| + |y_2-\hat{y}_2| + |y_3-\hat{y}_3| \cdots + |y_n-\hat{y}_n|}{n}$$

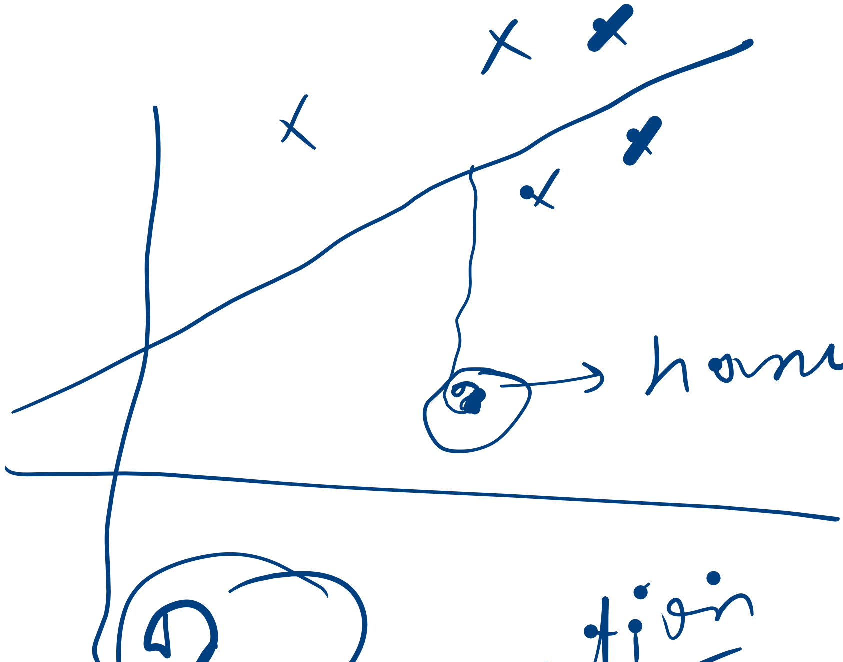
$$\boxed{mae = \frac{\sum_{i=1}^{n} |y_i-\hat{y}_i|}{n}}$$

Advantage

① Same unit

x | y pam LPA

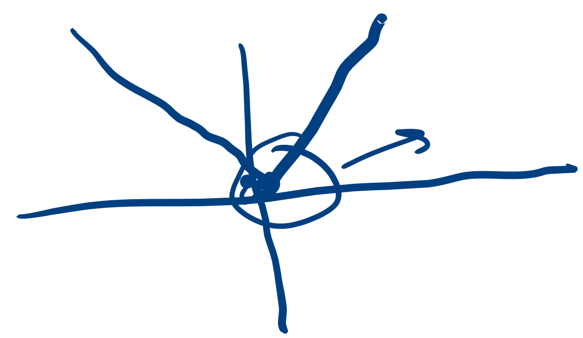
mae = 1.5 → LPA



→ handle outliers

② Robust outliers

Disadvantage MAE



③ $\frac{\partial}{\partial x}$ optimization

↓ cost function ← min or max

$$E = \sum \left(y_i - mx_i + b\right)^2$$

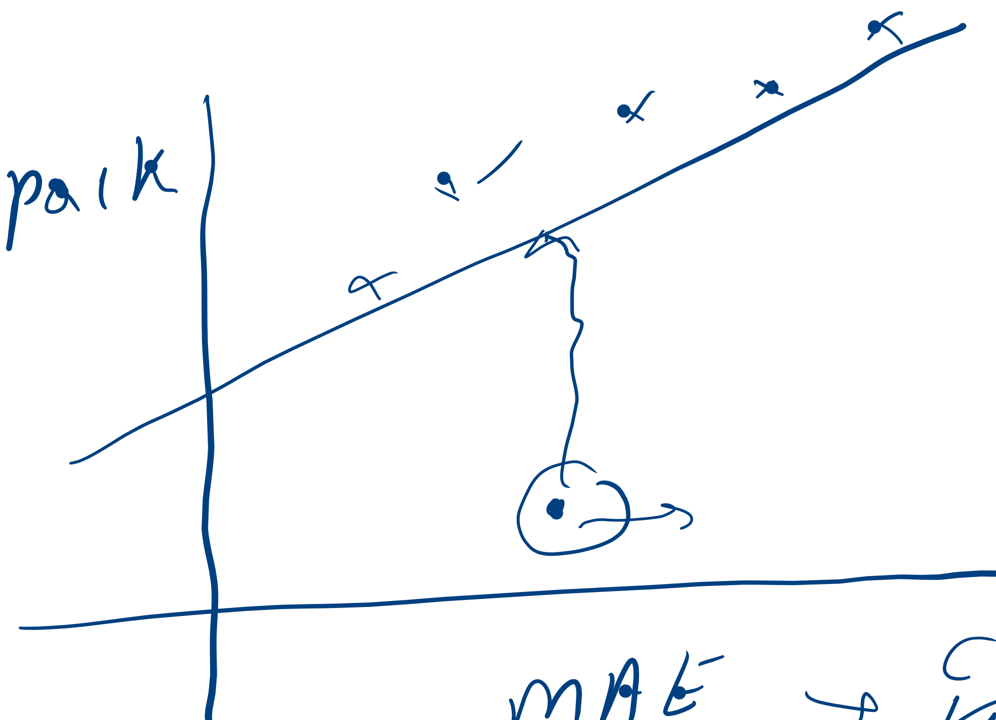
$$\bar{y}$$

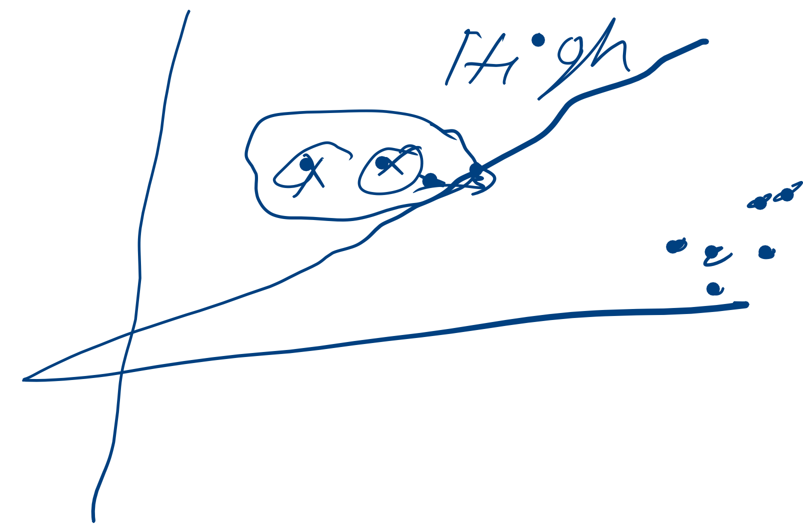
$$\frac{\partial E}{\partial m} = \frac{\partial}{\partial m} \sum \left(y_i - m x_i - \bar{y} + m\bar{x}\right)^2 = \sum m\left(x_i - \bar{x}\right)^2$$

$$(x_i - \bar{x})$$

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

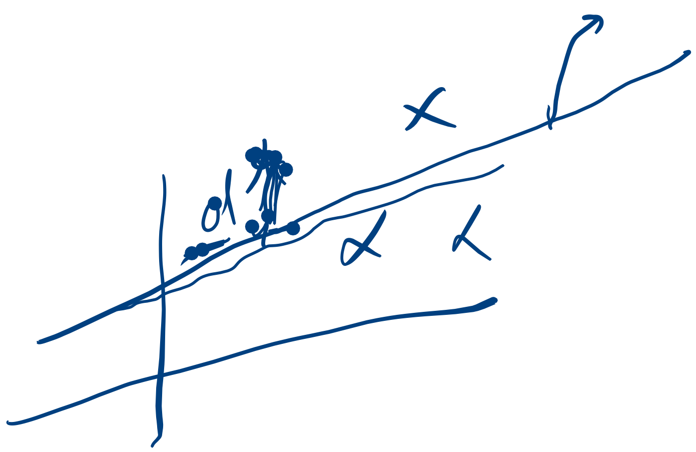
$$m = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

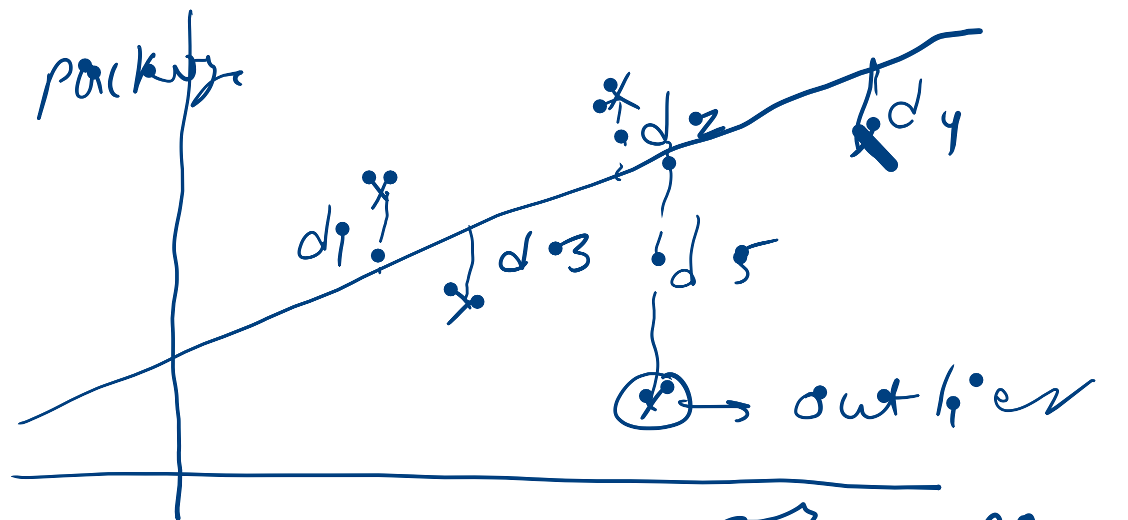




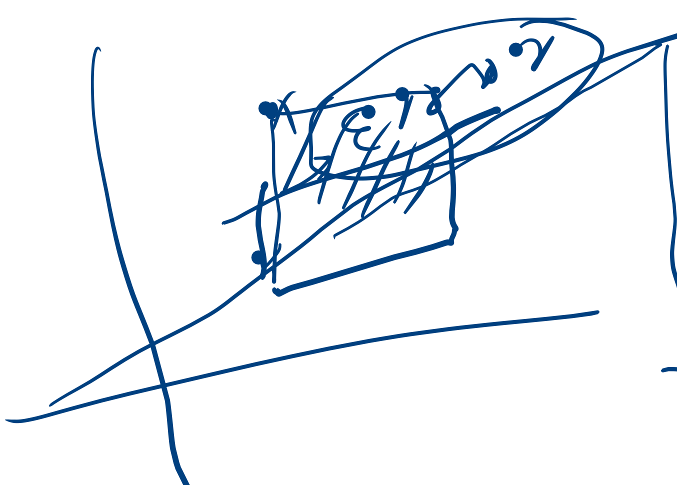
MAE → +ve → Sum

MSE → mean square error

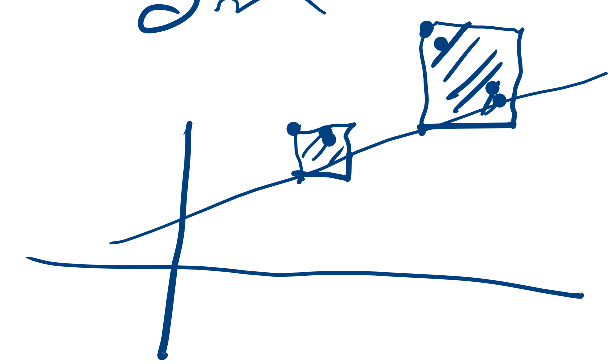




Loss funct



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



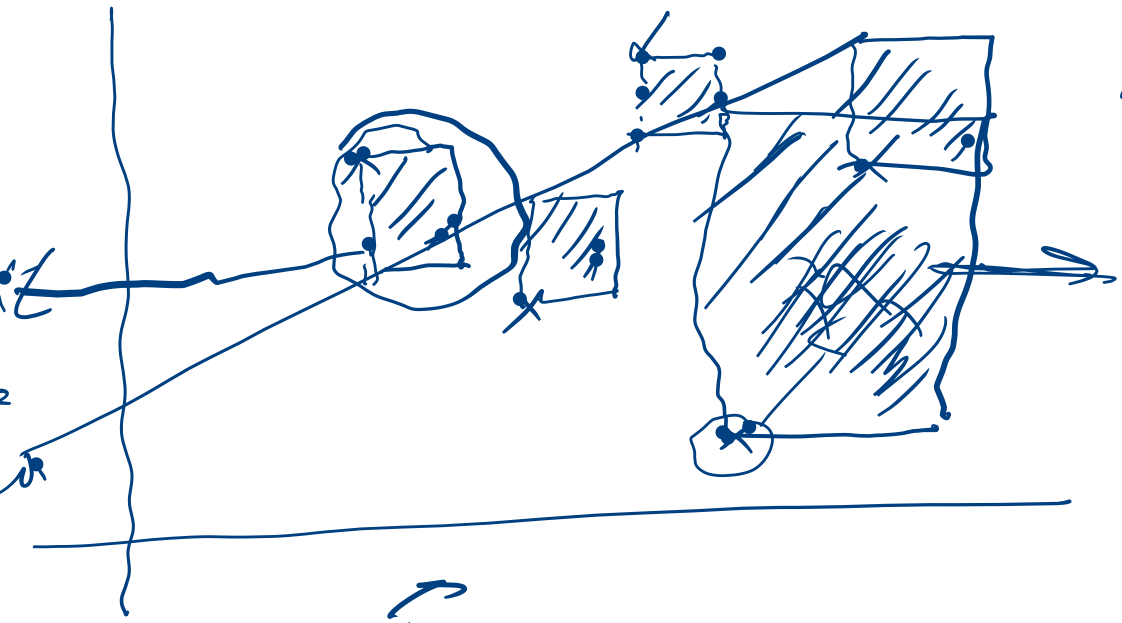
$m \times m = m^2$ area ← Square

MAE → (pa)

i) less outliers then we can MSE

more → MAE

Error reduction the area



MSE ↓ 11·25

Advantage

i) Loss function → differentiate · MSE

optima

Disadvantage

$\sqrt{11.25}$

$m \times m = m^2 \; ((pa)^2$

x	y
Output	Package

model-1

F2

model

0.98

e.g.

metrics

MSE

MAE

RMSE → Root mean square Error

Deep learning

$\sqrt{MSE}$

how much error.

advantage → ~~+~~ → unit

pichu $\sqrt{\cdot}$

$RMSE = \sqrt{1.5}$

$MSE = \sqrt{1.5}^2$

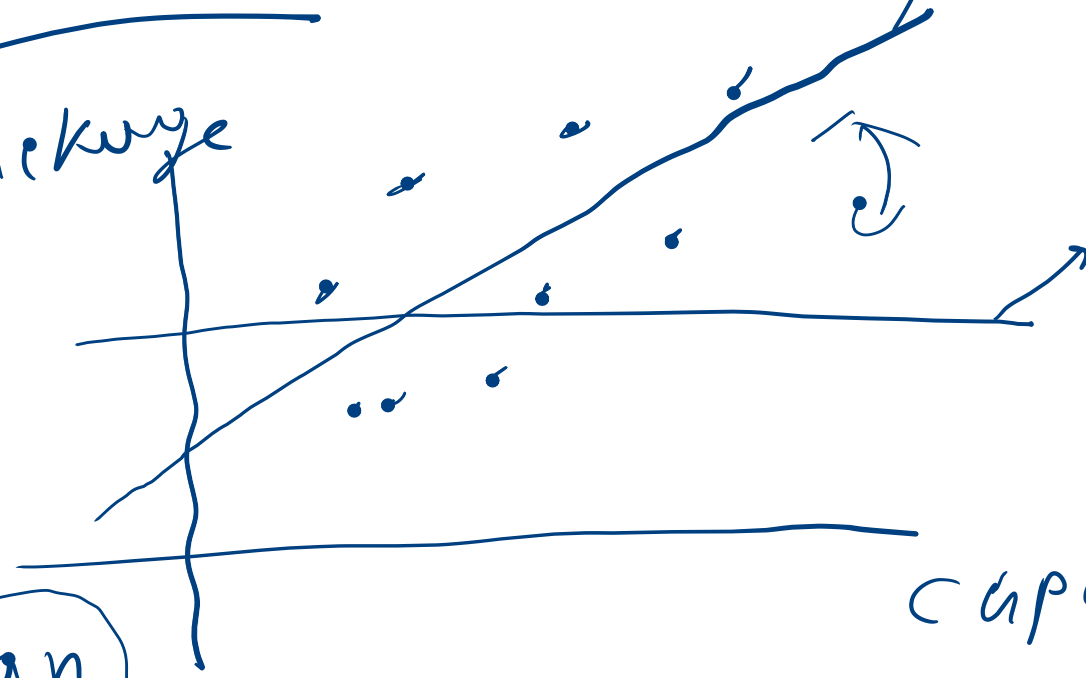
performance

R_2 Score

worst

CGPA	package

mean



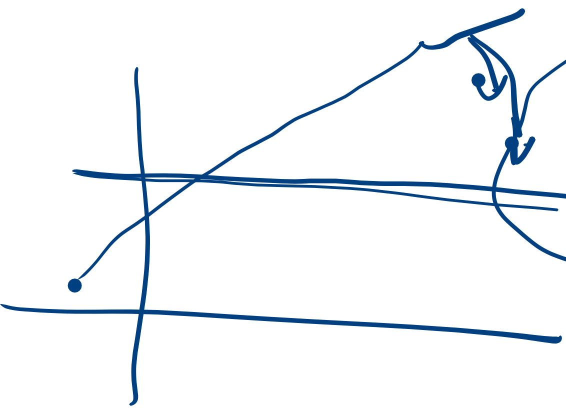
R_2 score → compare mean line with best line.

$$\boxed{R_2 = 1 - \frac{SSR}{SSm}}$$

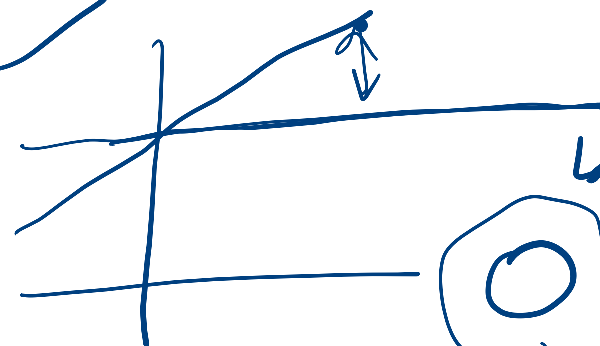
SS_R → Sum of square error regression line

SSm → Sum of square error of mean line

$$R_2 = 1 - \frac{SS_R}{SS_m} = 1 - \frac{\underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{R} \rightarrow 0}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2_m}$$



$R_2 = 0$



0

$R = m$

$R_2 = 1$

model perfect move → R_2 moves toward 1

$0 \le R_2 \rightarrow 1$ worst ~~move~~ move : R_2 moves to 0

$R_2 = 0.80 \rightarrow$ 80% of LPA variance due to CGPA

CGPA is partially

80%

input line

90%

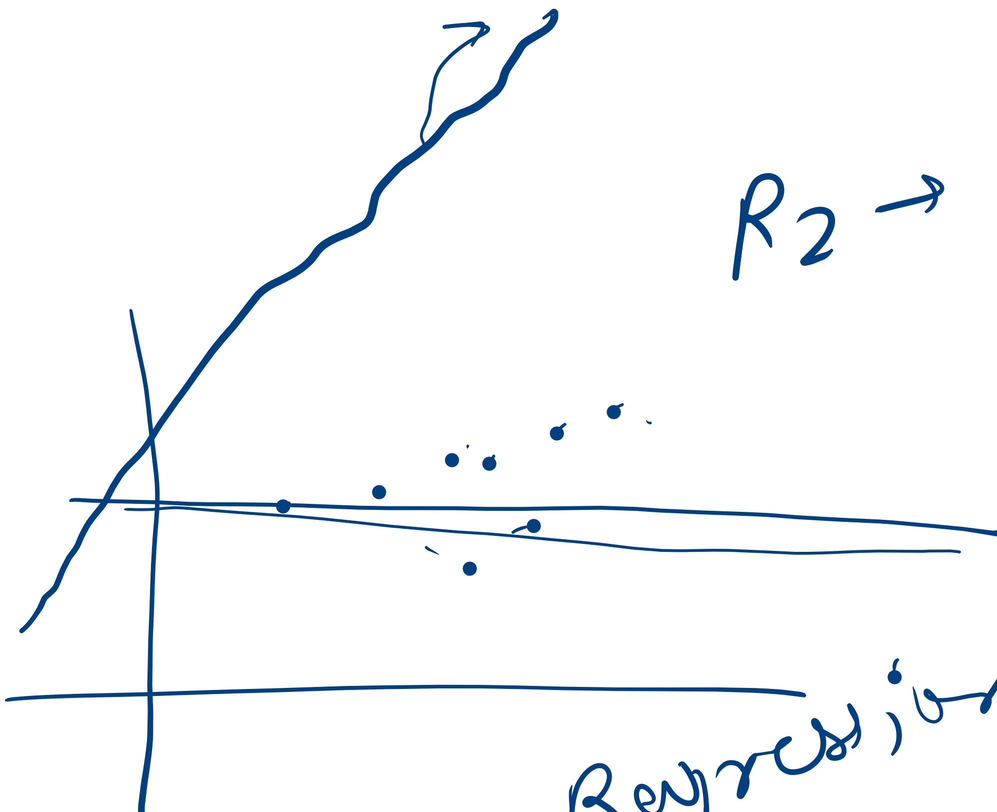
$R_2 \rightarrow$ negative

fonly capacity

day → capa

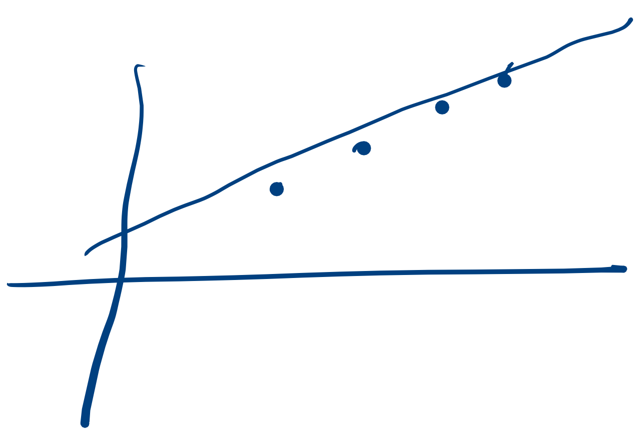
$R_2 \rightarrow$ negative

$= \dfrac{SSR}{SSm} > 1$



Regression line do more error the mean line

$\boxed{SSR > SSm}$



temp | cgpa | package

$R_2 = 80\%$ ↑ 90%

Adjust R_2 score

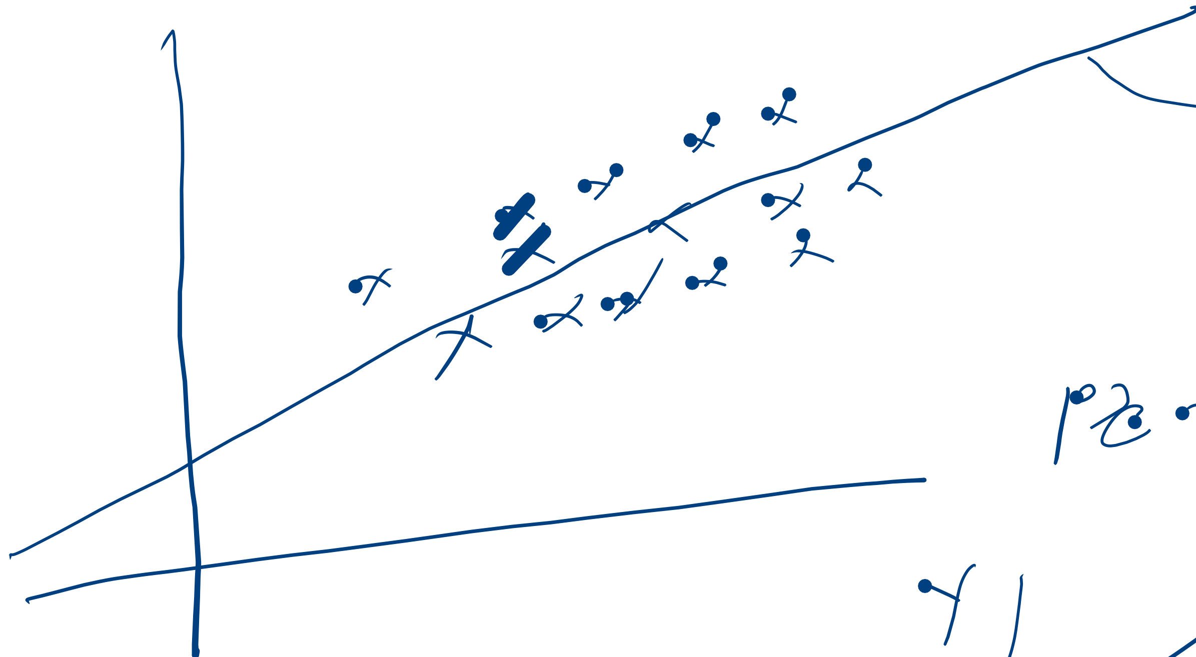
temp

1%

$$\text{Adjust } R_2 \text{ Score} = 1 - \left(\frac{(1 - R_2)(n-1)}{(n-1-k)}\right)$$

n = no. of rows

k = independent features.

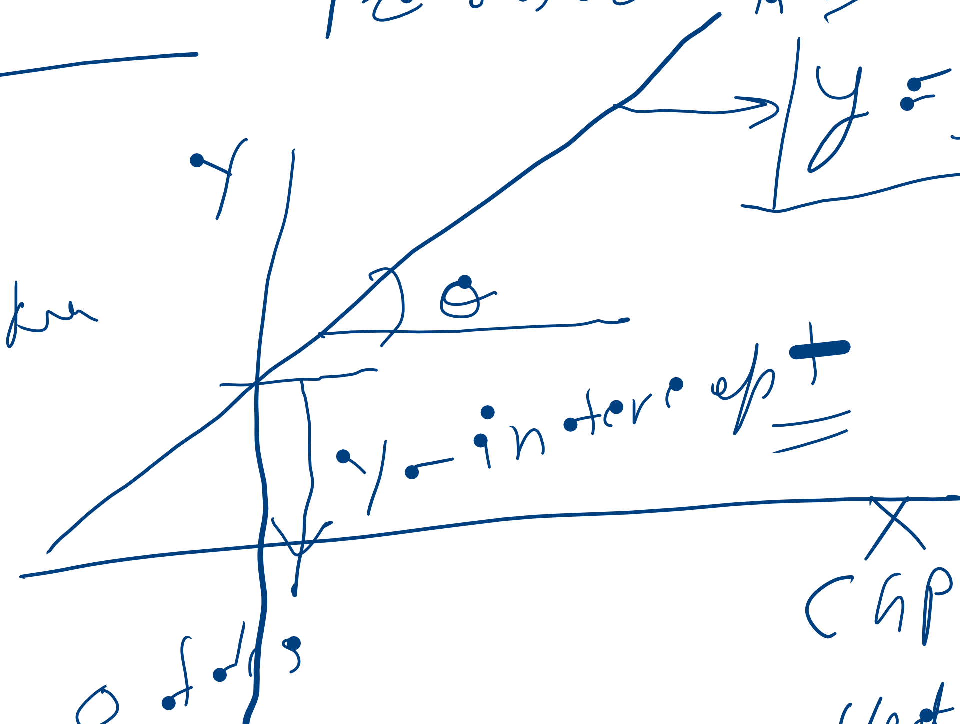


Best line → close to all the points

predict line

$$y = mx + b$$

slope



parku

$y = mx + b$ → capa

↳ parku

$b =$ offset

E	S
0	2L
1	

$b =$ offset.

$y = mx + b$

x → experience

y → salary.

$y = mx + b$

gradient

$x = \text{Exp} — 0$

$y = \text{salary}$

$y = mx$

$y = 0x$

m, b

$y = mx + b$