Introduction

Monday, May 10, 2021 3:53 PM

Linear Regression

Easy to understand

Regularization

LR → Supervised — Type

Regress

classi

unsup

Polynomial LR

Simple LR

Multiple LR

MLR

Simple

1 input col | 1 output

MLR

cgpa	package
6.66	3.01

cgpa

cgpa	gender	12th marks	state	package

Simple Linear Regression

Monday, May 10, 2021 3:54 PM

cgpa	package
7.1	3.5
4.7	1.2
8.9	4.2
8.1	3.9

$\boxed{\text{model}} \rightarrow$ package

cgpa

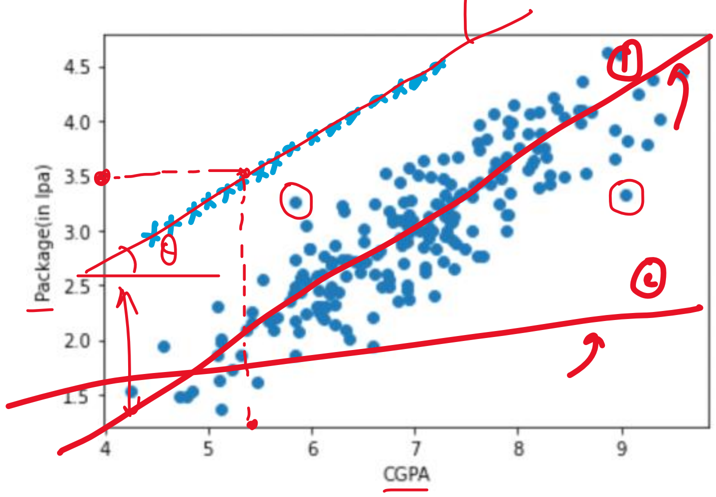
$\boxed{3.5\ \text{Lpa}}$

sort of linear

$\boxed{m, b}$

package $\leftarrow y = mx + b$, $x \rightarrow$ cgpa

200 students



$\boxed{\text{Why?}}$

Real word dataset

Best fit line

{ m → slope
b − y intercept }

LR

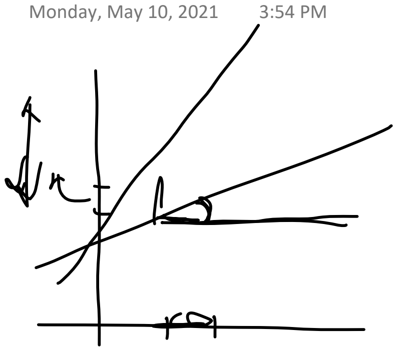
→ $\boxed{m, b}$

Code Example

Monday, May 10, 2021 3:54 PM

Intuition

Monday, May 10, 2021 3:54 PM



$y = mx + b$

$\boxed{package = m \times cgpa + b}$ → b

$\boxed{b = 0}$

m / b

$package = m \times cgpa$

exp → 0

m → weightage

m ↓ ↑

b

package

exp | package

offset

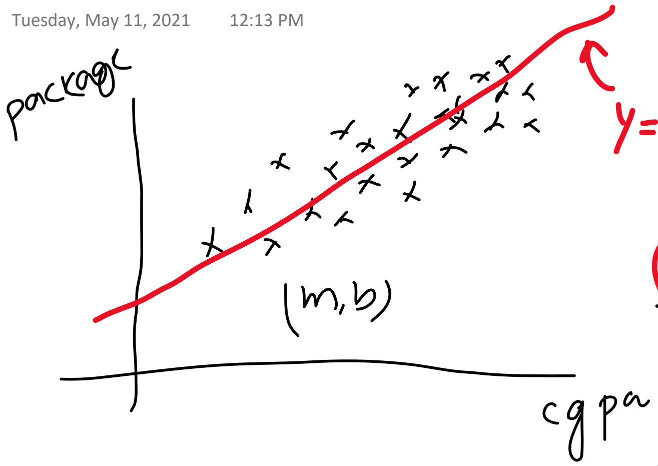
$package = \boxed{m \times exp} + \boxed{b} = 0$

$\boxed{p = 0}$

$p = m \times exp$

How to find m and b?

Tuesday, May 11, 2021 12:13 PM



$y = mx + b$

(m, b)

Higher 1-D → LinearReg → SubRegression

- Closed form solution → direct formula → OLS
- Non-closed form → Gradient Descent

$y \to$ pack

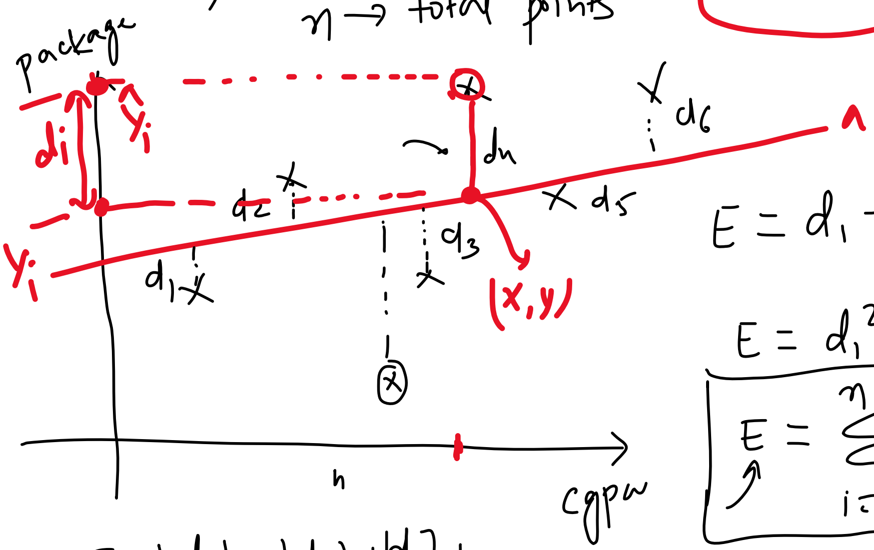
$x \to$ cgpa

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

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

$\bar{x}, \bar{y} \to$ mean (m, b)

$n \to$ total points



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

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

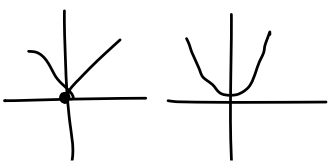
$$\boxed{E = \sum_{i=1}^{n} d_i^2}$$ ← Error function (J)

(m, b)

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

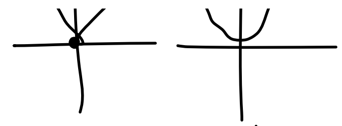
R1

R2 →



$d_i = (y_i - \hat{y}_i)$

R2 →



(m, b)

$d_i = (y_i - \hat{y}_i)$

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

(m, b)

$\hat{y}_i$ $mx_i + b$

(m, b)

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

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

$E(m, b)$

b ↑ $m = 0$ →

$y = f(x)$

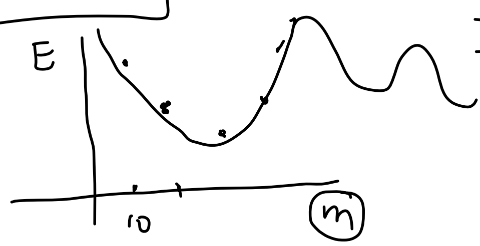
$$E(m, b) = \boxed{\sum_{i=1}^{n} (y_i - m x_i - b)^2}$$

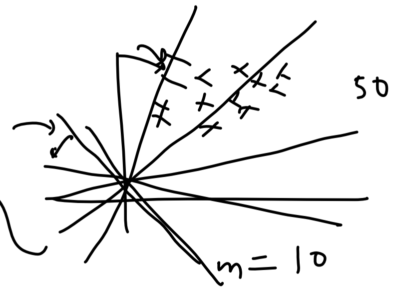
$b = 0$

minimum

(m, b)

$y = f(x)$



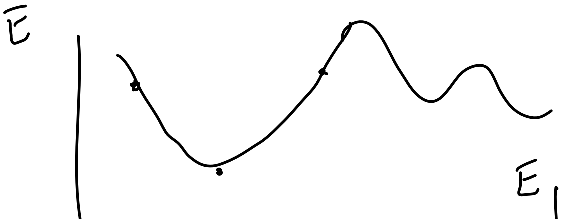


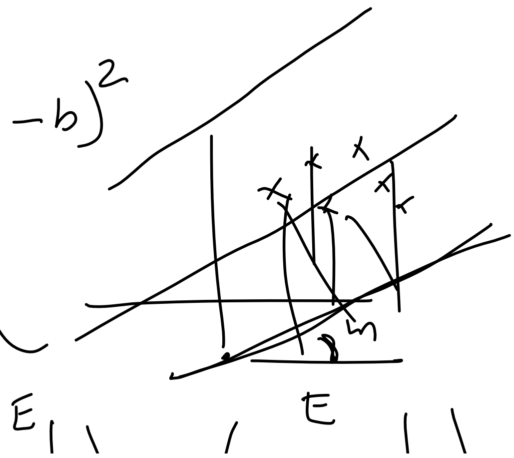
$\boxed{b = 0}$

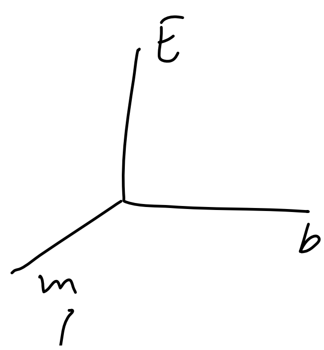
$m = 1$

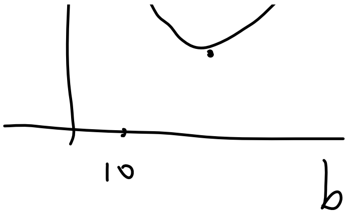
$$E(m) = \sum_{i=1}^{n} (y_i - m x_i)^2$$

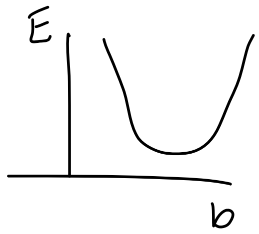
$$E(b) = \sum_{i=1}^{n} (y_i - x_i - b)^2$$

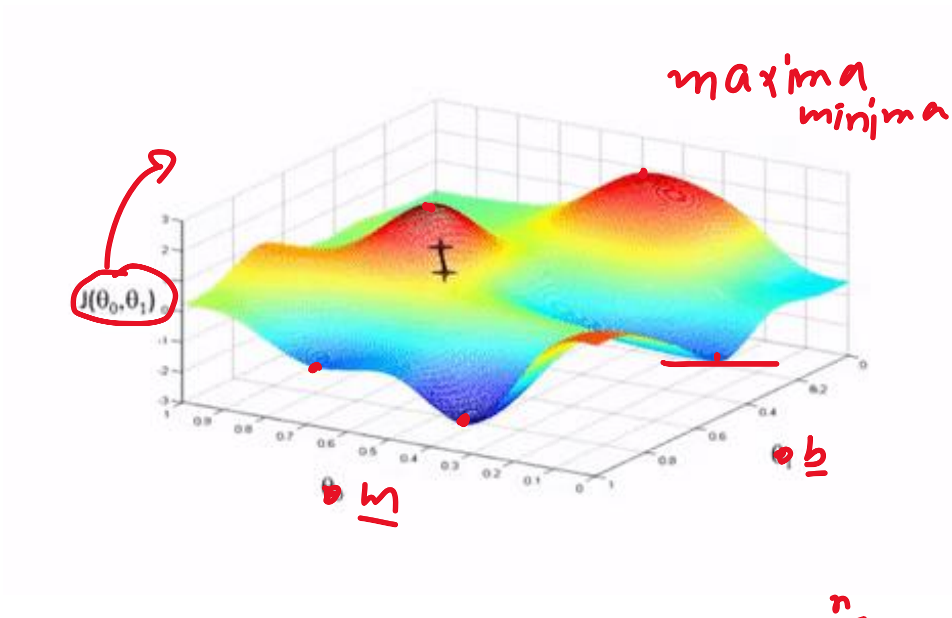












$\rightarrow E(x)$

$$\frac{dE}{dx} = 0$$

$f(x, y)$

$$\frac{\partial E}{\partial m} = 0, \quad \frac{\partial E}{\partial b} = 0$$

(m, b) n

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

$$= \sum \frac{\partial}{\partial b} (y_i - m x_i - b)^2 = 0$$

-1

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

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

$$\underbrace{b + b + b + b + \dots + b}_{n \text{ time}} = nb$$

$$\frac{\sum_{i=1}^{n} y_i}{n}$$

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

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

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

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

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

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

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

$$= \sum -2\,(y_1 - mx_i - \bar{y} + m\bar{x})(x_i - \bar{x}) = 0$$

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

$$= \sum \left[(y_i - \bar{y}) - m(x_i - \bar{x})\right](x_i - \bar{x}) = 0$$

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

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

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

Code from scratch

Tuesday, May 11, 2021 12:14 PM

Regression Metrics

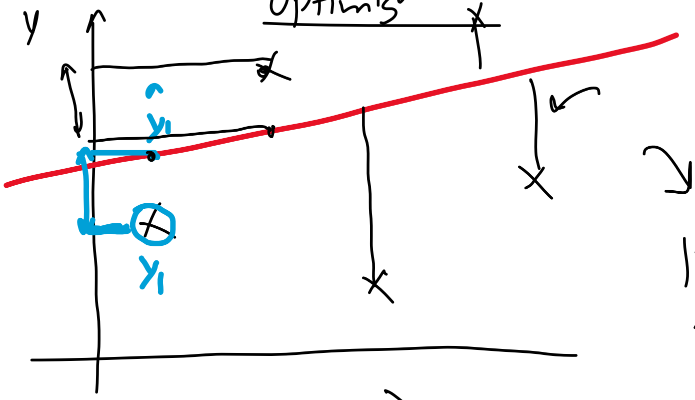
Thursday, May 13, 2021 11:56 AM

1) MAE
2) MSE
3) RMSE
4) R2 score
5) Adjusted R2 sore

MAE

Thursday, May 13, 2021 11:56 AM

Optimization cgpa $\xrightarrow{x}$ | y package (lpa)



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

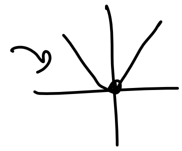
$$\boxed{mae = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}}$$

1.5 lpa

Advantage

1) same unit
2) Robust outliers

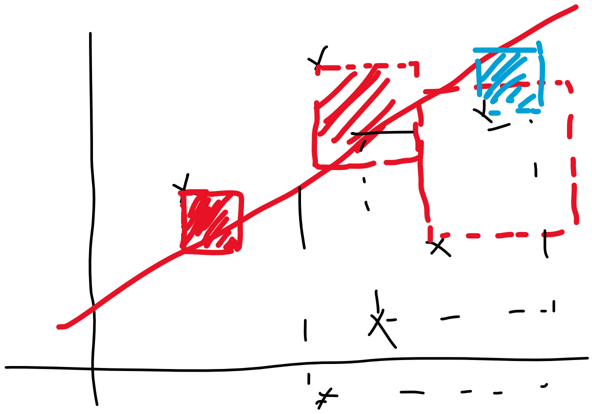
Disadvantage



MSE

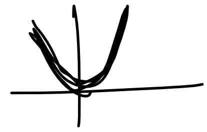
Thursday, May 13, 2021 11:56 AM

mean squared error cgpa | y pack (lpa)



11.25

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



mse

L function

$(y_1-\hat{y}_1)^2$

Advantage

L function

Disadvantage

Robust to outliers

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

y_lpa

mse – $(lpa)^2$

RMSE

Thursday, May 13, 2021 11:56 AM

$Rmse = \sqrt{mse}$

mse

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

rsme → loss functions

0.72

R2 score

y

rmse - lpa

y - lpa

benefit

disadvantage

Robust outliers

1.5

1.5 lpa

mse mae

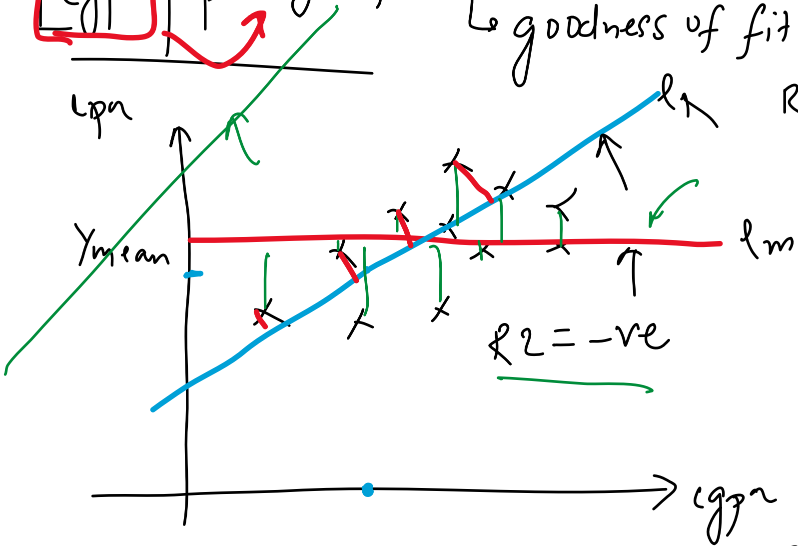
R2 Score

Thursday, May 13, 2021 11:56 AM

$\boxed{100} \rightarrow$ mean ↳ 3.4

$\boxed{\text{cgpa}}$ | package (lpa)

Coeff of determ
↳ goodness of fit



$R2 = -ve$

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

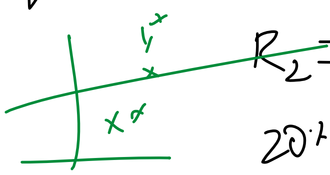
$$R2 = 1 - \frac{\left[\sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right]_{Reg}}{\left[\sum_{i=1}^{n} (y_i - \bar{y}_i)^2\right]_{m}}$$

$= 1$ $= 0$

↳ $\rightarrow 0$

$R_2 = 0$

Sum of squared error Reg



$R_2 = 1$

20%

$0 \rightarrow R2 \rightarrow 1$

$\boxed{SSR > SSm}$

80%

$\boxed{R2 \rightarrow 0.80}$ $\rightarrow$ $\boxed{\text{cgpa}}$ explain 80% of variance in lpa

cgpa | lpa ↑ 20%

cgpa | iq | lpa $\quad \boxed{\text{lpa}}$

Adjusted R2 score

Thursday, May 13, 2021 11:57 AM

0.80 0.90

R2 score

cgpa | lpa

iq

temp

$R_2 \uparrow$

Adjusted R2

$[1 - \uparrow]$ $\downarrow$ $[1 - \downarrow] \uparrow$

R2 →

n → no. of rows

K = independent

K=1, K=2, K=3

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

Multiple LR

temp

Adj R2 ↓

iq

$R^2_{adjs} \uparrow$