x_1

$x_1 \quad x_2$

$x_1, \quad x_2 \quad x_3$

$x_1 \quad x_2 \quad x_3 \quad x_4$

Polynomial

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0$$

Poly = multiple.

linear = $mx + b = 1$

quadratic = $m_1 x_1^2 + m_2 x_2 + b = 2$

Cubic = $m_1 x_1^3 + m_2 x_2^2 + m_3 x_3 + b = 3$

$= m_1 x_1^4 + m_2 x_2^3 + m_3 x_3^2 + m_4 x_2 + b = 4$

1. Fit, Fit_transform, transform
2. Git Intro
3. split strategy.

1. Fit, Fit_transform, transform

standard scaler

$$\frac{X-\mu}{\sigma}$$

X = input
μ = mean
σ = Standard deviation

Height (Independent)	Weight	Age → Target
100 Samples		

Scale.fit() → find μ, σ

Scale.fit_transform() → find μ, σ and transform $\frac{X-\mu}{\sigma}$

100 samples → Train, Test

70 samples → Train

30 samples → Test

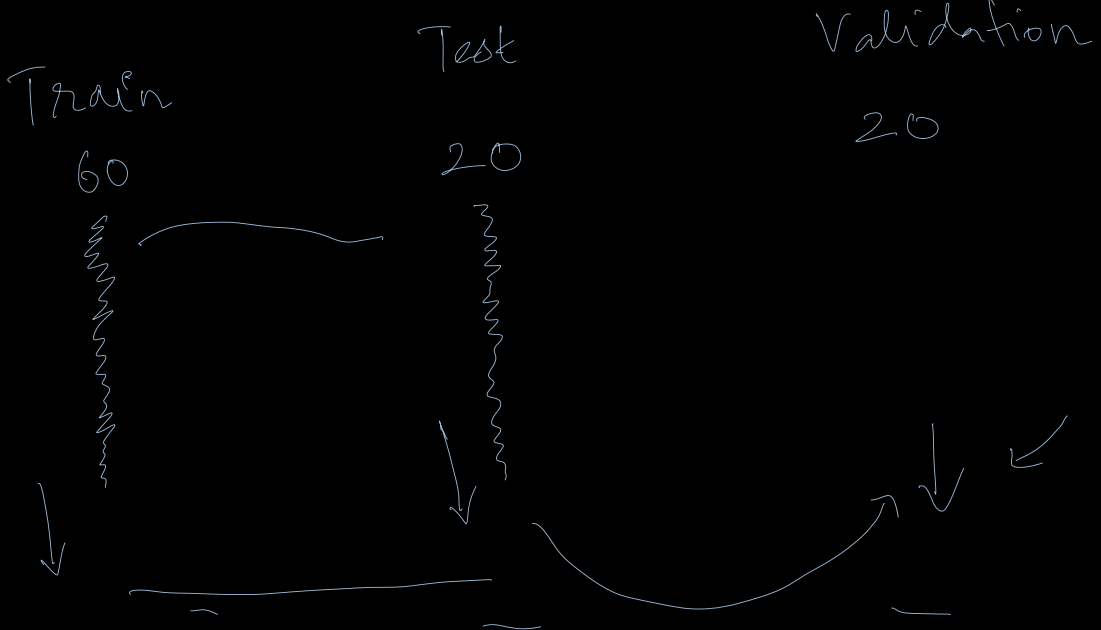
Train → Scale. fit_transform() → μ, σ, transform $\frac{x-\mu}{\sigma}$

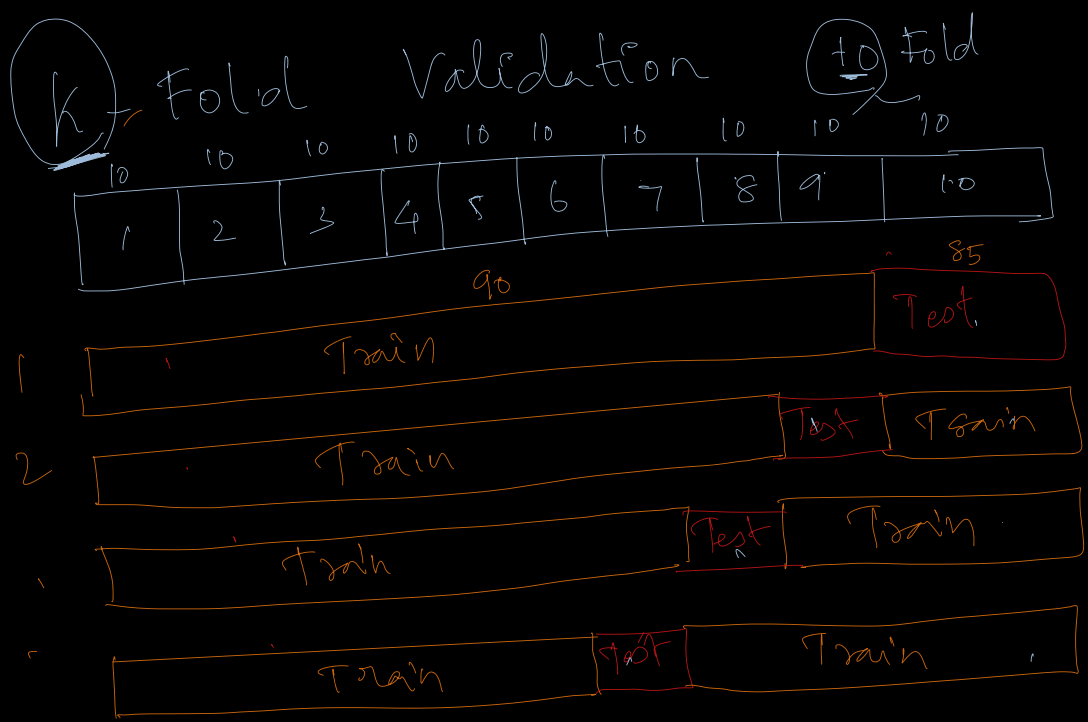
never fit Test → Scale. transform() ← transform

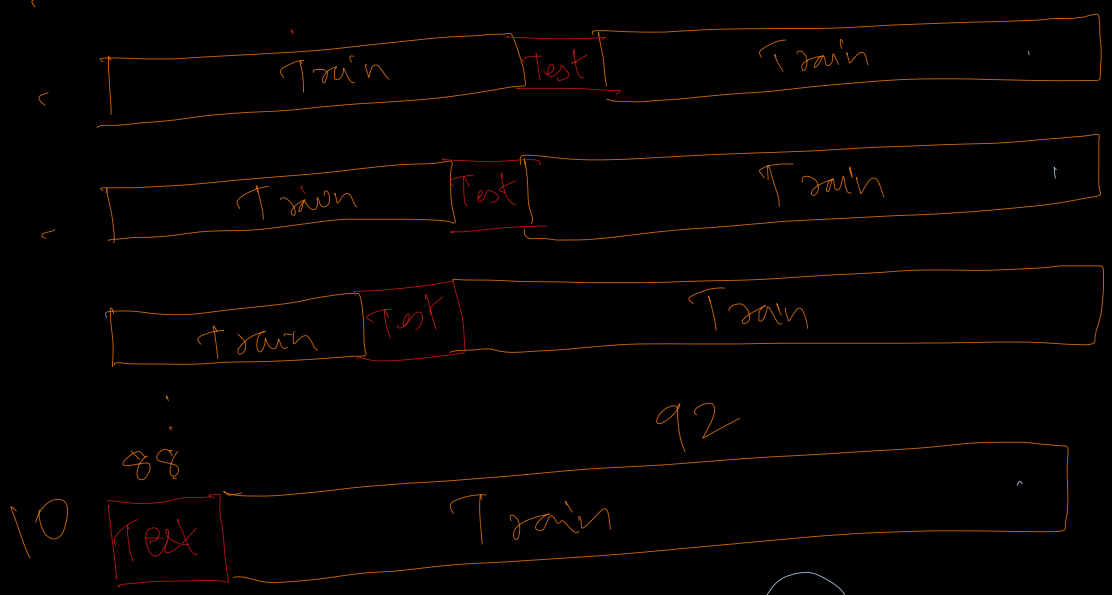
One data

2. Split strategy

Train	Test
60	40
70	30
75	25







10 Train R2 = 80
10 Test R2 = 82

10 Test

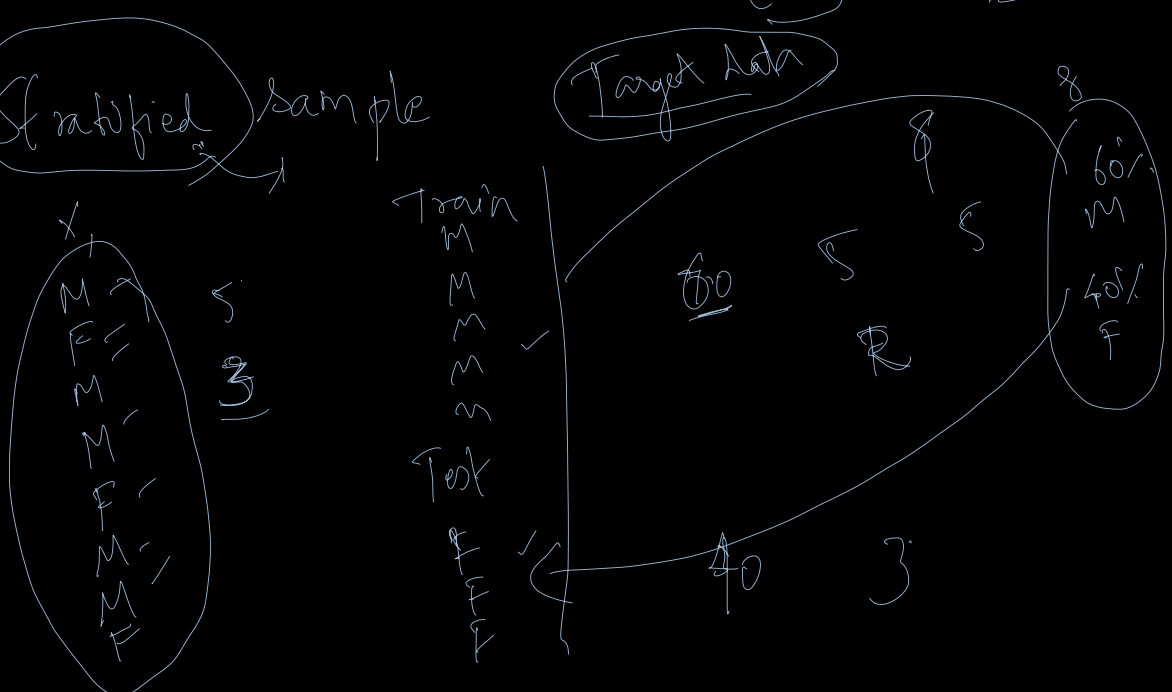
standard deviation

high → Inconsist
low → Consistent

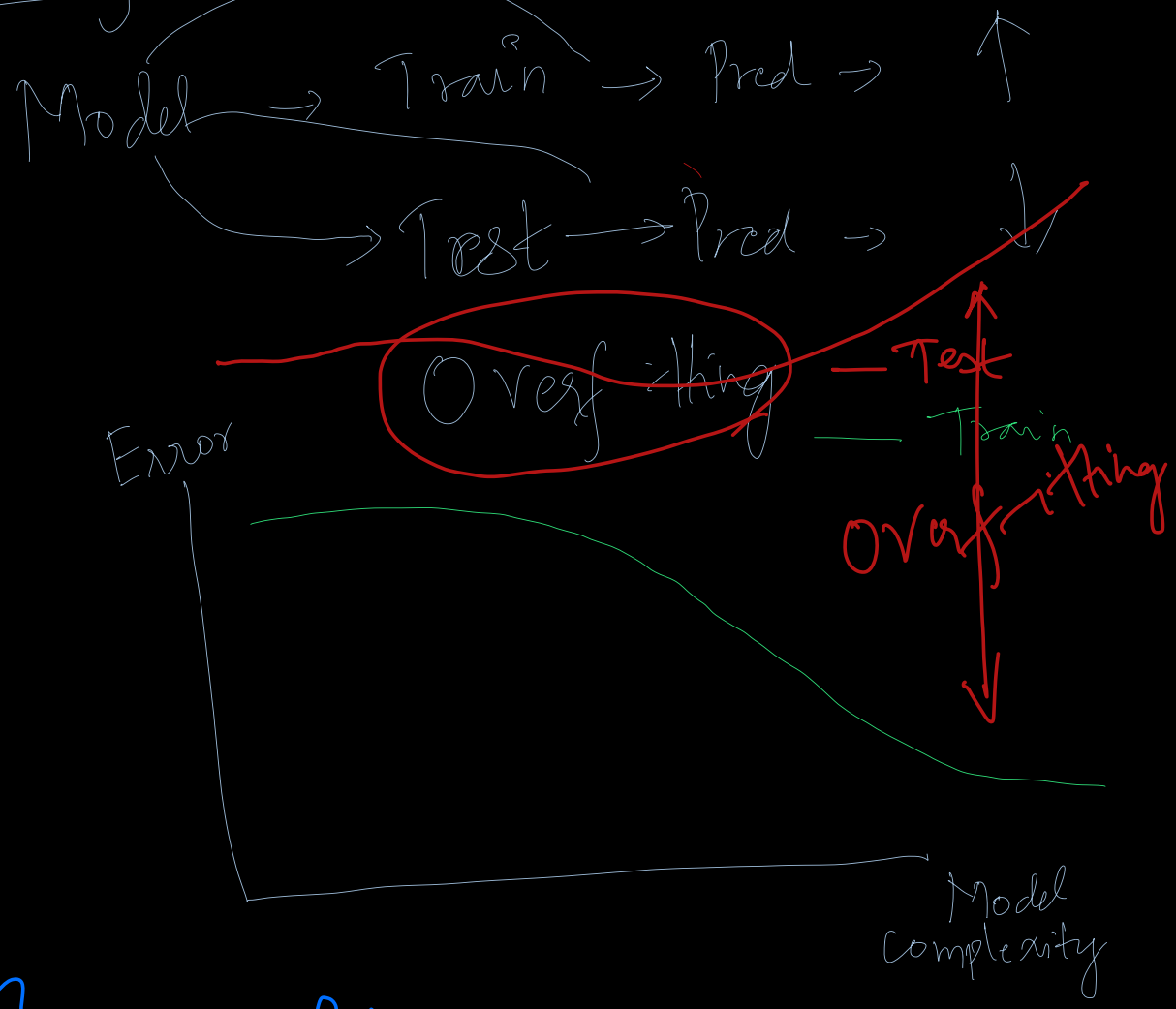
Model

Stratified sample

Target Data



Overfitting:



Generalization

Student learning

5 year q/A → 100
500

Mugged this 500 quest Answers

500/500 100 Overfitty

6 year

Concept → 60/100

Student

500

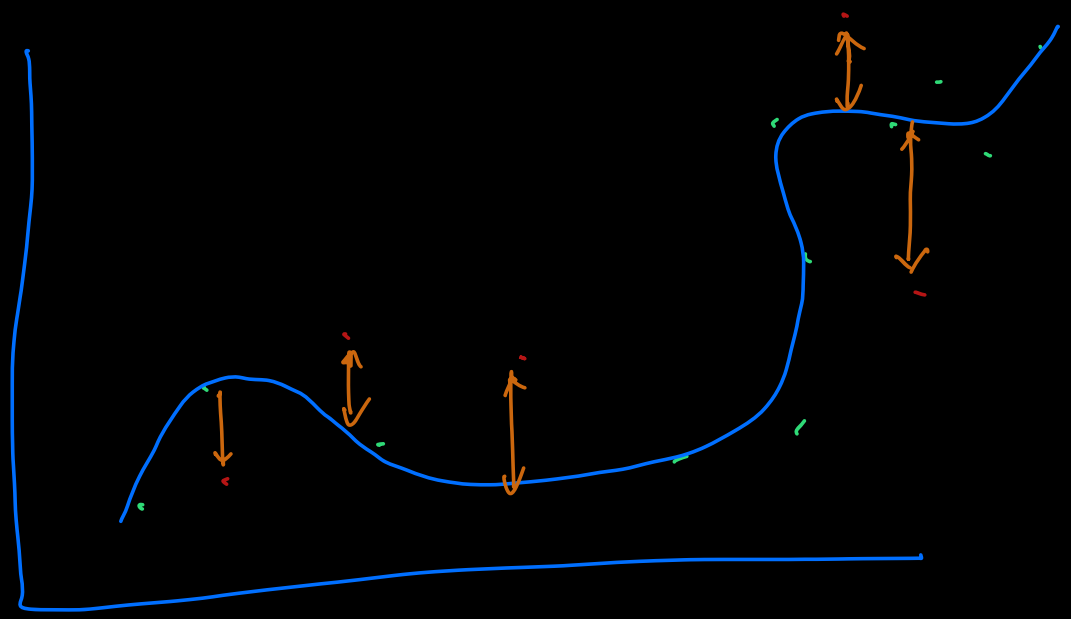
300 Concept

Generalisation

500

300 500 70%

Actual learning 65%

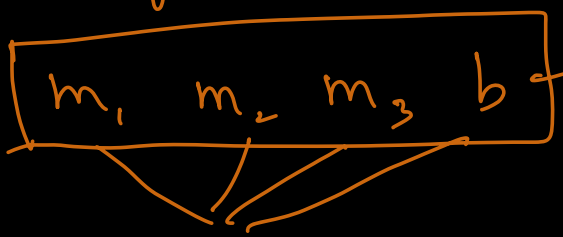


Regularization

1. l_1 (Lasso)
2. l_2 (Ridge)
3. ElasticNet

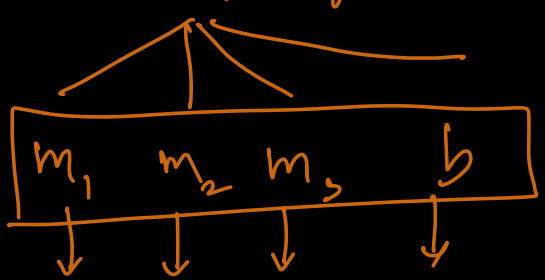
Regularization

$$y = m_1 x_1 + m_2 x_2 + m_3 x_3 + b$$



→ finding

Overfitting :

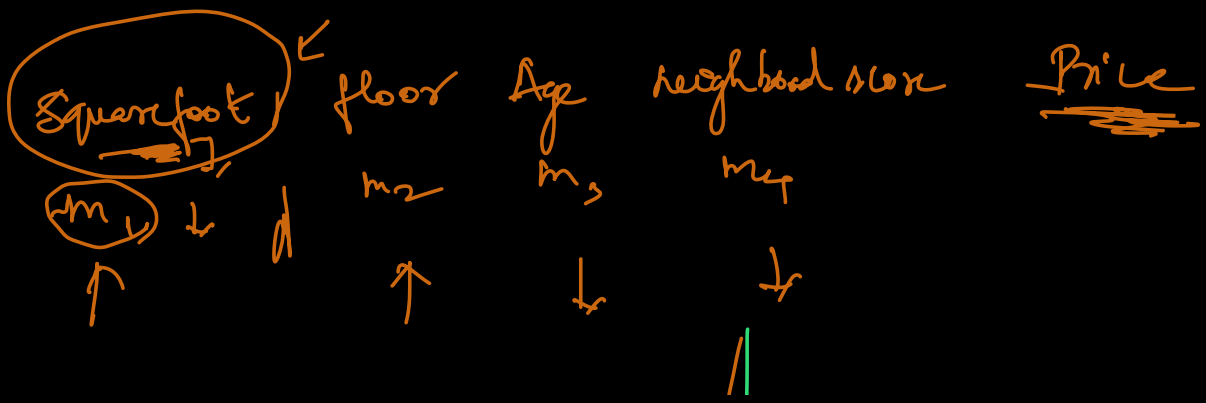


→ Generalized

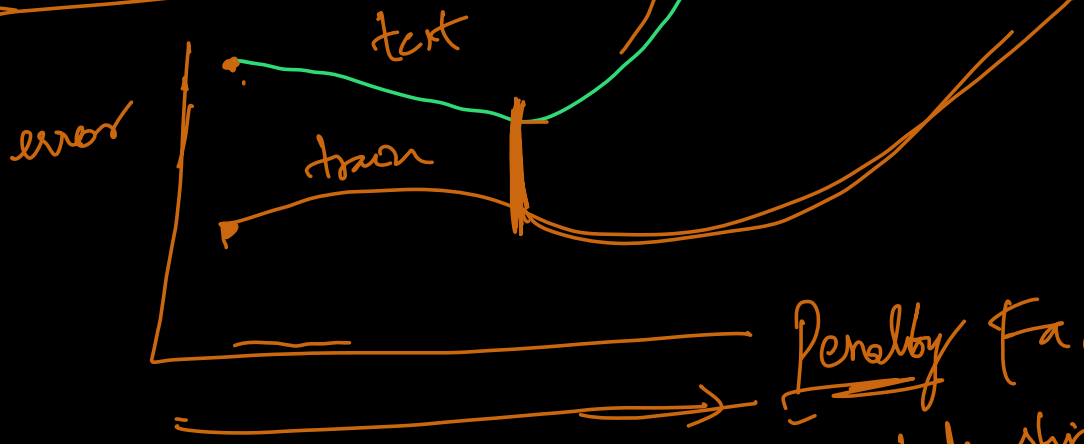
regularly making check of not making itself to be a huge number

error ↓

m_1, m_2, m_3, b ↓



Penalty on Coefficients Size



Lasso Regression (L1) least Absolute shrinkage and selection operator

Criteria

↓ MSE

Goal

Slope and Intercept

Criteria

↓ MSE and ↓ slope and intercept

$$\downarrow E = \sum (y - \hat{y})^2 \downarrow$$

$$\downarrow E = \sum (y - (mx + b))^2 \downarrow$$

Partial derivatives.

$$E' = \left[\sum (y - (m'x + b'))^2\right] + \lambda(m' + b')$$

Penalty factor

Lasso

$$= \sum y - (w_1 x + w_0))^2 + \lambda |w_1 + w_0|$$

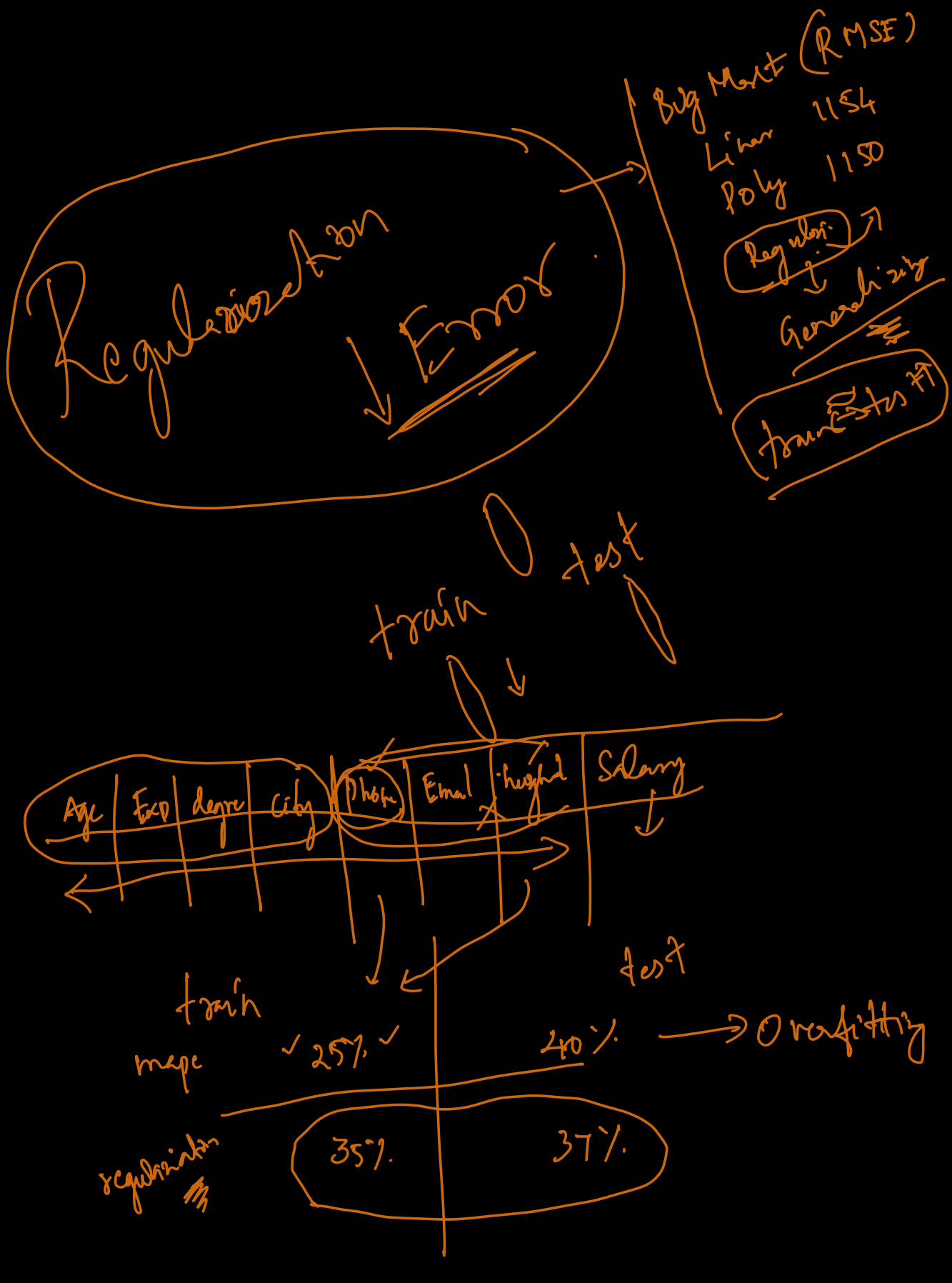
Lasso will reduce all the coefficient and make the coefficient zero if it is not important

Lasso → Overfitting → feature selection

Lasso

Room → Old stuff → 90%

Strict Mom / Spouse → clean → throwing unimportant stuff


Regularization
↓ Error
Big Mart (RMSE)
Linear 1154
Poly 1150
Regularization
↓
Generalizing
train
test
Age
Exp
degree
city
Phone
Email
Weight
Salary
train
test
mape
25%
40%
overfitting
regularization
35%
37%

2. Ridge Regression L2

$$E = \underbrace{\sum (y - (\hat{m}x + \hat{b}))^2}_{\text{Ridge}} + \lambda \underbrace{(m^2 + b^2)}_{\text{Penality factor}}$$

Ridge → reduce the coefficient to good extent but it will never make any coefficient zero

3. Elastic Net : L1 + L2

$$E = \underbrace{\sum (y - (\hat{m}x + \hat{b}))^2}_{\text{Ridge}} + \underbrace{\lambda_1 (m^2 + b^2)}_{\text{Penality factor}} + \underbrace{\lambda_2 (m + b)}_{\text{Lasso}}$$