17/10 Linear Regression

1. Assumptions of Linear Reg
2. Simple Linear Regression
3. Simple Equation
4. Linear Algebra –
5. Gradient descent
6. Error Metrics
7. Data Split
8. Practical

Agenda

1. Practical
2. ~~Hackathon~~ Data
3. Overfitting and Underfiting
4. Variance and Bias Tradeoff
5. Complete the Practical Assignment
   . Polynomial Regression
6. Lasso Regression
7. Ridge Regression
8. Elastic Net Regression
9. Deployement
10. Cloud (AWS, GCP)
11. Practical

Simple Linear Regression

$$y = mx + b$$

Multiple Linear Regression

$$y = m_1 x_1 + m_2 x_2 \cdots\cdots + b$$

1. Over fitting

Train → Model → apply → Test apply

Metrics

230 | 240

90% Revision | 30% Semester

250 | 150

Train error is less
Test error is more

Overfitting → Make the model generic

Underfitting | Make the model complex

Train error is high
Test error is very high

Exam
Student

80% | 20%

90% ← 135% → 85%

40% | 10%

degree-1

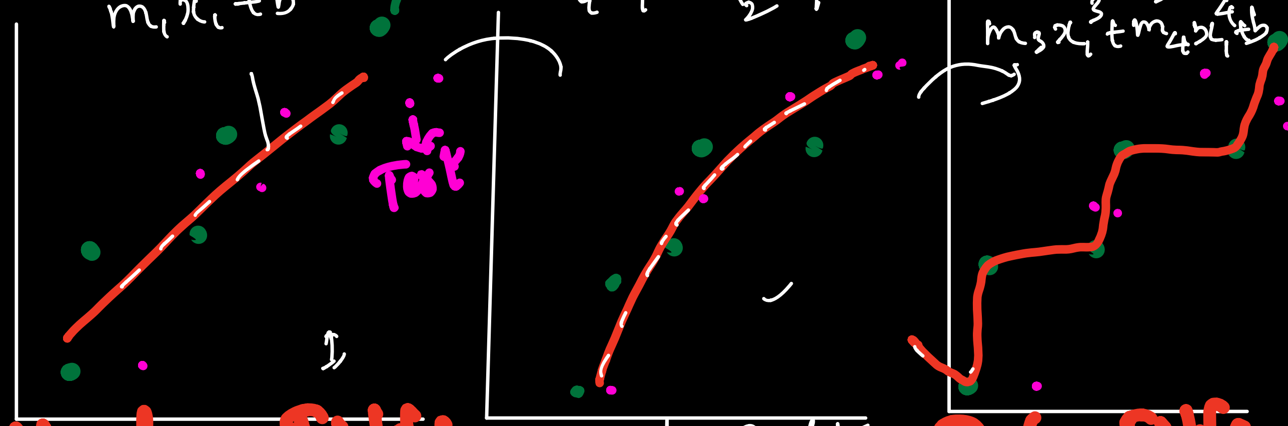
m_1x_1+b

degree 2

$m_1x_1+m_2x_1^2+b$

degree 4

$m_1x_1+m_2x_1^2+m_3x_1^3+m_4x_1^4+b$



Under Fitting — Just Right — **Over Fitting**

Problem

Train → Error ↑
Test → Error ↑

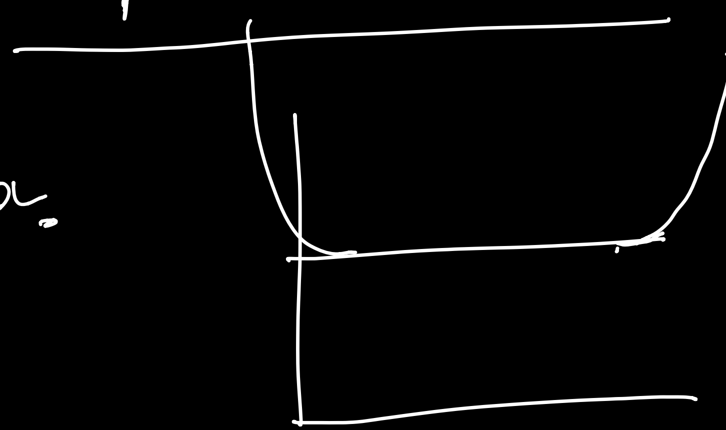
Train Error ↓
Test Error ↓

→ Train Error ↓
Test Error ↑

Polynomial

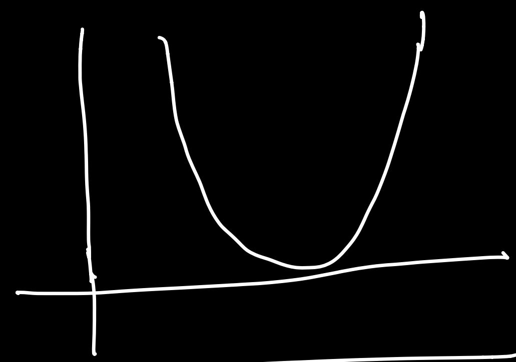
$$y = mx^1 + b$$

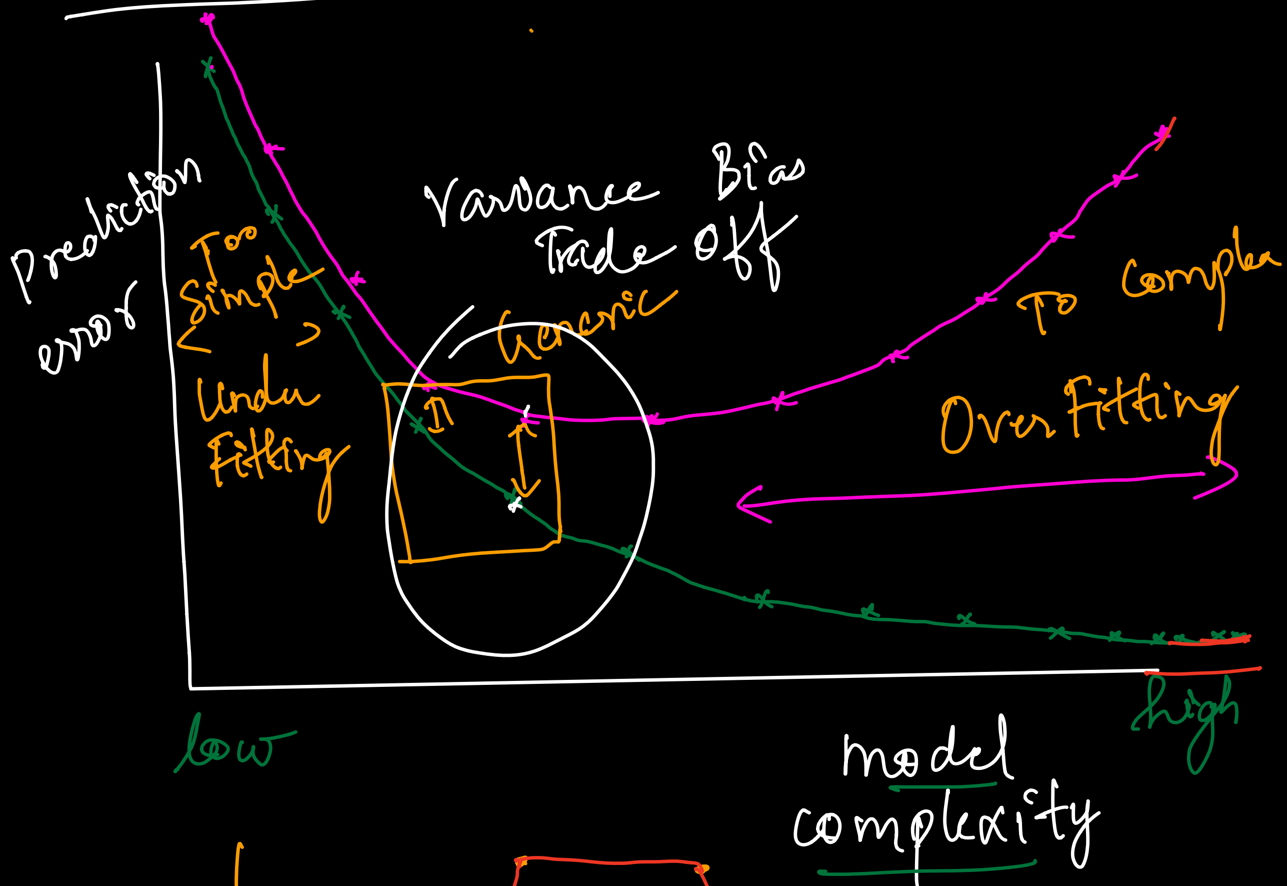
$$y = m_1x_1^1 + m_2x_1^2 + m_3x_1^3 + b$$



$$y = x$$

$y = x^2$



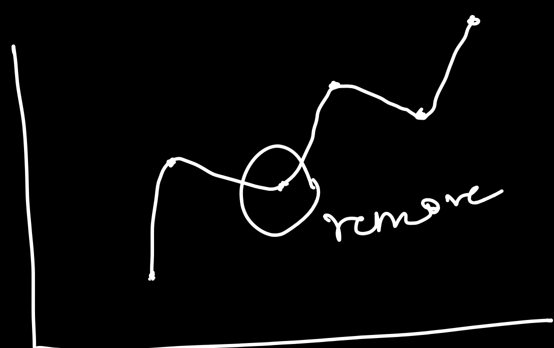

Variance Bias
Trade off
Prediction error
Too Simple
Under Fitting
Generic
To Complex
Over Fitting
low
high
model complexity


~0
~100

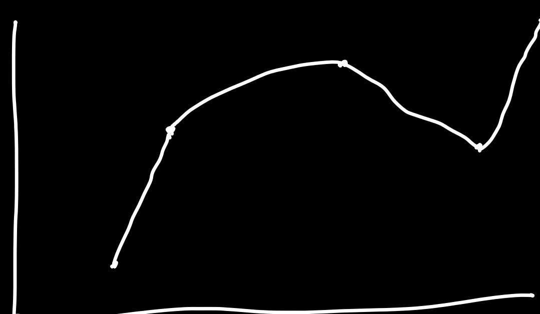
High Variance	High Bias
1. Over Fitting	Under Fitting
2. Model Complex	Model is very simple

High Variance

1. Over Fitting
2. Model Complex



3. Small change in Data



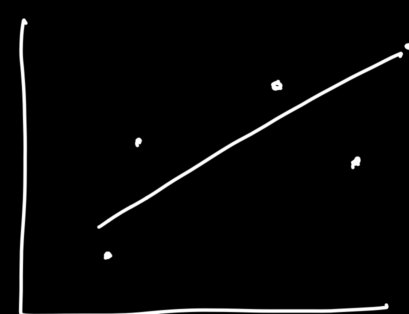
there is going to be a big change in co-efficient

High Bias

Under Fitting

Model is very simple

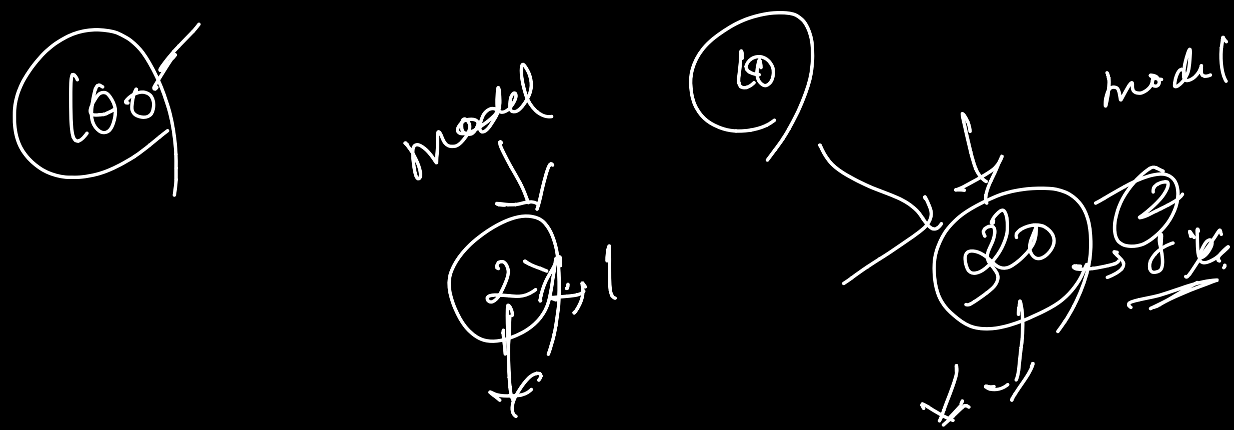
Small change in Data

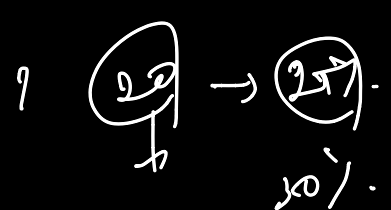


There is not going to be a big change in co-efficient

Ideal Model should have

Low Variance Low Bias





1000

110

Overfitting — Common

Regularization — Solution Technique

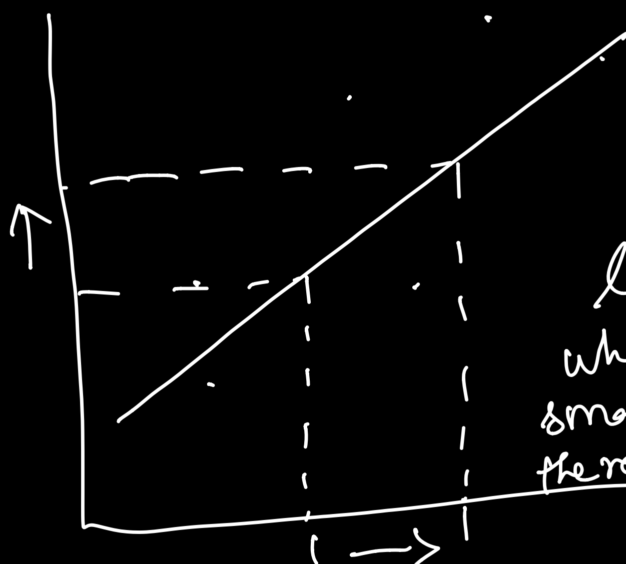
1. Ridge Regression (L2)
2. Lasso Regression (L1)
3. Elastic Net

Cost function in Linear Regression

MSE

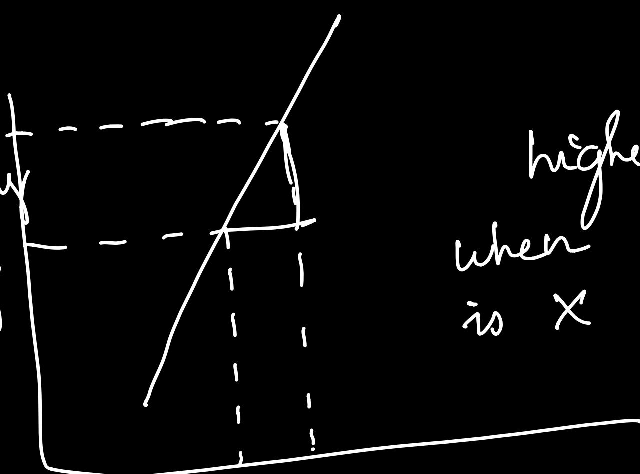
$$\frac{\sum (y-\hat{y})^2}{n}$$

$$\boxed{\frac{\sum (y-(\hat{m}x+\hat{b}))^2}{n}} \downarrow$$



lesser slope
when there is a small change in x
there is a small
change in y

there is
a possibility
of overfitting



higher slope
when small change
is x big change y

L2 Regularization (Ridge):

$$\text{cost} = \frac{\Sigma(y-\hat{y})^2}{n} \quad LR$$

$$\boxed{\frac{\Sigma\left(y-(\beta_0+\beta_1 x)\right)^2}{n}}$$

Cost

$$\frac{\sum (y - (\beta_0 + \beta_1 x))^2}{n} + \sum (\beta_i)^2$$

$b + mx$

$\beta_0 = 500$

$\beta_1 = 500$

$$\frac{\sum (y - \hat{y})^2}{n} = 1000 \quad LR$$

error

Ridge Regress $= 1000 + (500^2 + 500^2)$

$= 1000 + 250000 + 250000$

$= 501000$

partial derivative

$mx + b$

$$\boxed{\beta_1 x + \beta_0}$$

Linear Regression

cost function = $\dfrac{\Sigma(y-(\beta_1 x+\beta_0))^2}{n}$

agenda → find β_0 and β_1 (min)

Constraint → reduce cost function

---

when slope is higher

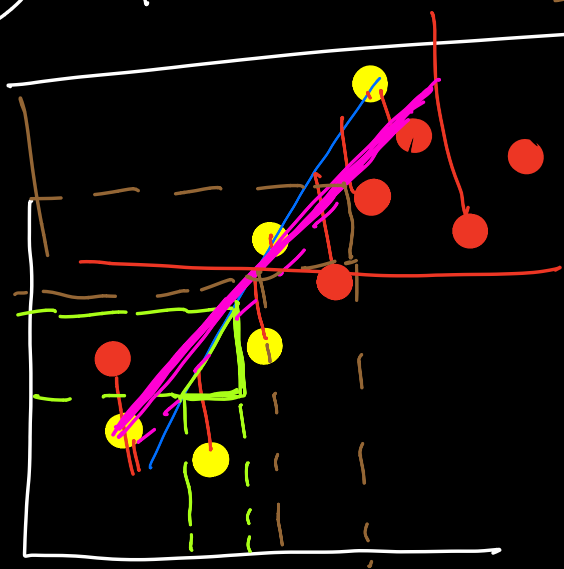
small rate of change in X

will have higher rate of change in Y

→ There is a possibility of Overfitting when you have higher slope

---

↳ Reduce the Co-efficient



Ridge

This increases a small in the train Data
↓
To match the error with Test data

Ridge Regression

1. Agenda:
   Find the β_0 β_1

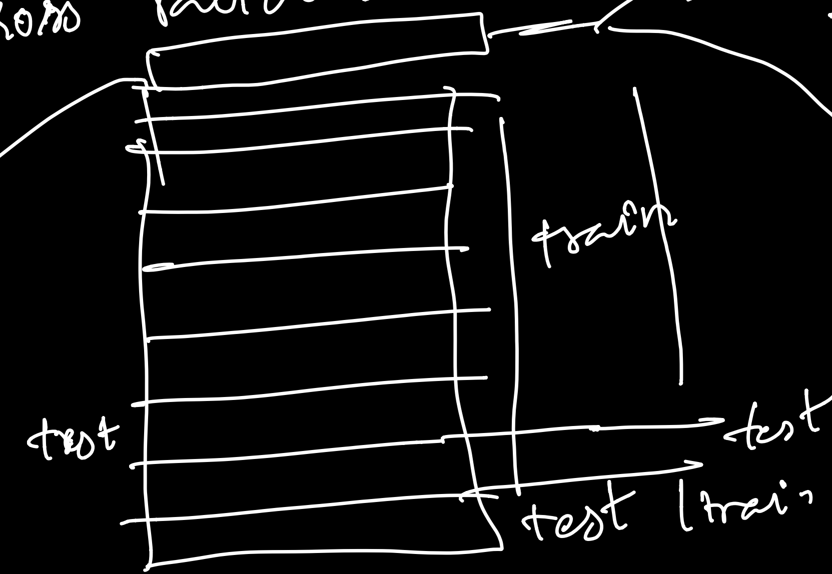
2. Contraint
   | reduce the Cost function
   | reduce the Coefficient

Cost function for Ridge ✓

$$\sum \frac{\left(y-(\beta_0+\beta_1 x)\right)^2}{n} + \sum_{i=0}^{n} (\beta_i)^2$$

Penalization

Cross validation → 800 RMSE



10 times → Standard deviation

50 ✓

1100 RMSE

1050 - 1150

Benchmark

Cost function for Ridge ✓

$$\sum \frac{\left(y - (\beta_0 + \beta_1 x)\right)^2}{n} + \lambda \sum_{i=0}^{n} (\beta_i)^2$$

Penalization

Hyper parameter $\rightarrow$ adjustment

$0 \rightarrow \infty$

Linear Regression

my slope be near to zero

Tuning

	RMSE Train	Test
$\lambda = 0$ =		
$\lambda = 0.1$ =		
$\lambda = 0.2$ =		
$\lambda = 1$ =		
$\lambda = 100$ =		

2. Lasso Regression (L1)

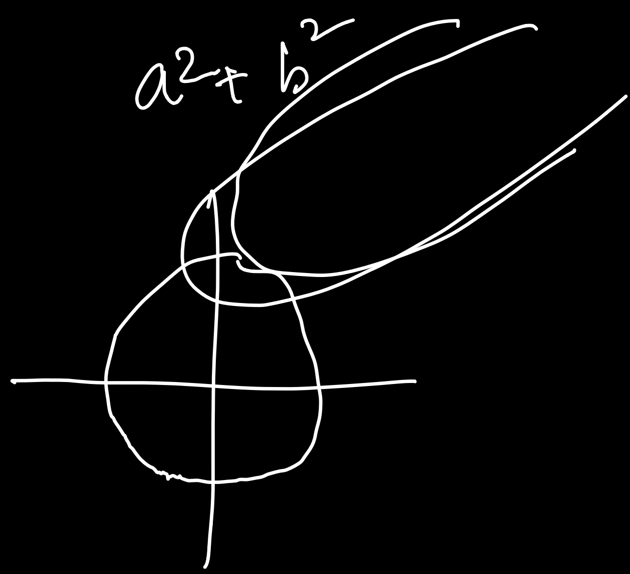
Cost function for Lasso

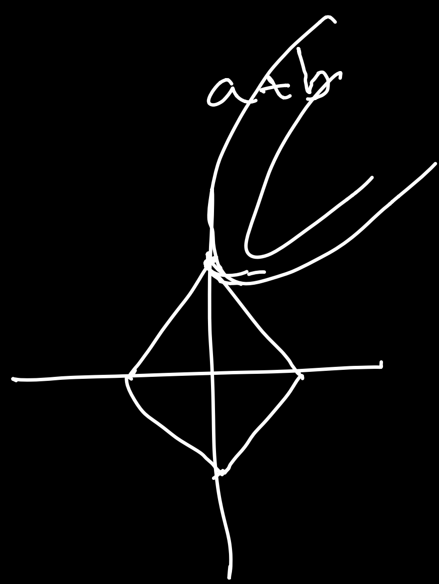
$$\frac{\sum \left(y - (\beta_0 + \beta_1 x)\right)^2}{n} + \lambda \sum_{i=0}^{n} |\beta_i|$$

↓

Penalization

<u>Feautre Elimination</u> = 0





3. Elastic Net

Lasso + Ridge

Cost function for Elastic Net

$$\sum \frac{(y-(\beta_0+\beta_1 x))^2}{n} + \underbrace{\alpha}_{\text{Alpha}} \underbrace{\left(\sum_{i=0}^{n} (\beta_i)^2 + \lambda_2 \sum |\beta_i|\right)}_{\text{Penalization}}$$