Cross validation, variance bias, regularisation

L

(1) What was the differ US Price &
Bigmart?
→ AnalyticsV. House Price

(2) How are they differ?

→ Dataset
→ Train → 'y' → Train
→ Test
→ Test → No 'y'

→ Dataset
→ Train → Train
→ Test

Cross - Validation

100%

25
25
25
25

75% Training → set

(1)

RMSE - 25.

25% Testing Set

→ One testing

(2)

→ Testing

Training

RMSE = 30.056

(3)

→ Testing

(4)

→ Test

→ Training

Lr = LinearRegn ()

→ How will be the actual est ??

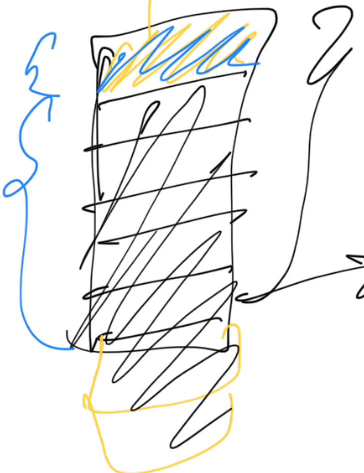
4-fold Cross Validation

4th data

4 → Quaters of the data → Training,

Leave One out → Test

→ Testing



10 → 10 fold Cross Validation

→ 9 - Training

→ Validation of Performance of our model In real life

---

4 - fold Cross Valid

① Train RMSE = 45
Test RMSE = 47

② Train RMSE = 45
Test RMSE = 7[illegible]

③ Train RMSE = 50
Test RMSE = 92

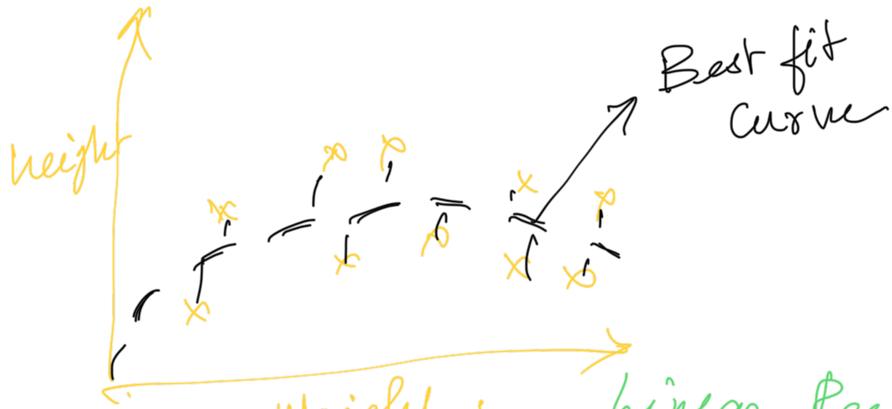
Train = 40
Test = 89

Overfitting & Underfitting

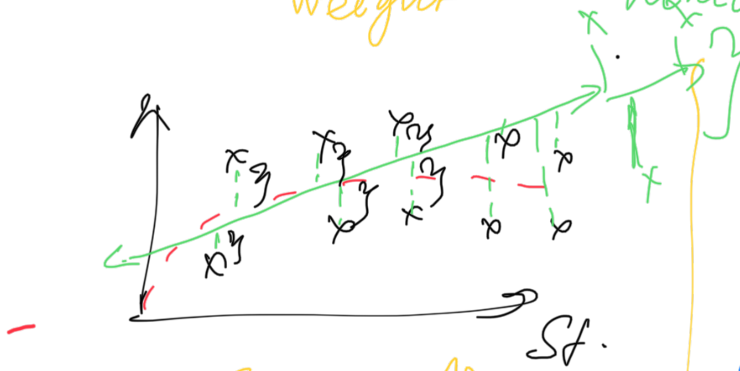
	Train	Test
Overfitting —	Performs too good	Poor Perform
Underfitting —	Performs not good	Perform not good

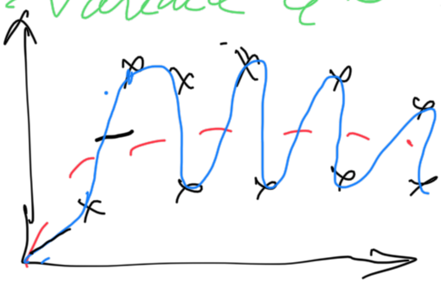
Variance & Bias Trade off

↳ Terminology Ate



Linear Reg. → Variance & bias





Least Square line
(Or)
Regn line
↓
St. line.

Good fit   great fit

High Bias
↓
higher bias
highly simple

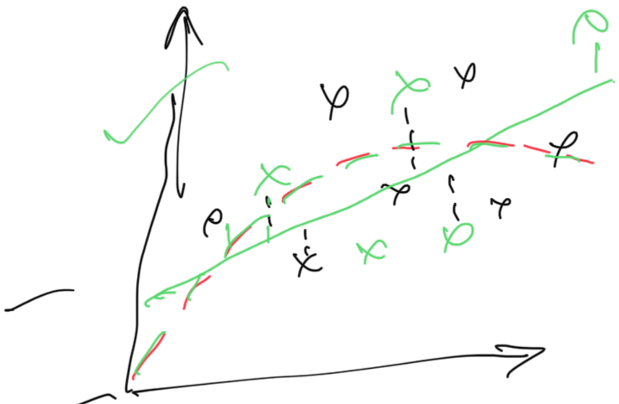
Squiggly line

This line can fit arcs so well. w

Curved line has very low bias

Complex model

lower bias



→ featur m.

→ Variance with featur is less

Variance with feature value is more

Over

Overfitting on Training set → Low Bias & High variance

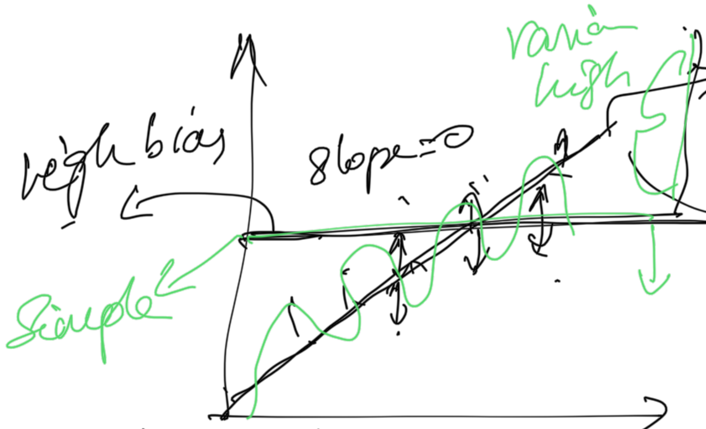
great fit → high low bias
low variance } → Sweet Spot

Regression → high bias
high v... } Underfit

Bias → Simplicity } → Trade off
Variance → Complex

Too simple | Too complex

Bias → Simplicity ... variance



bias is the model lin Eq.

linear Reg.

Optimal bias
Optimal variance } good fit.

Models	Bias (Simplicity)	Variance (Complexity)
Overfitting →	low	high.
Underfitting →	high	low
Perfect fit →	low	low

→ Comp

Vari

Variance
↓
S.D ⇒ variance
σ } variance
σ

Comp
bias – Simplicity
Vari → Complexity

S.D

Variance – Variability

~~Vari~~ Overfitting

Train	Test
almost 0	high

~~Variance~~

Complexity

$y = mx + c$

M

$y = mx + c + 2 \cdots$

$\cancel{y = mx + c}$

Rival

~~Deviat~~ Underfitting

Train	Test
High	High.

complex ~~Simple~~

Simple / ~~complex~~.

$y = mx + c$

$m = 0$

mean line

Model Complexity

Train	Test
Medium to low	Medium to low

Variance

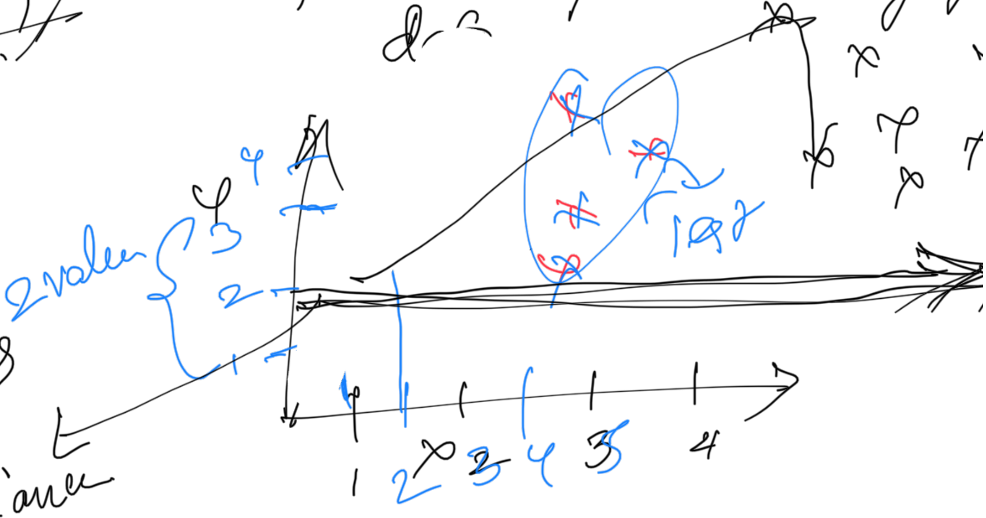
$\frac{\sum (y - \hat{y})^2}{n}$

x low high

Bias → low complexity

Variance → high Complexity

Sensitivity to underlying data



High Bias low variance

Variance as Sensitivity towards

Chaps.

Bias → low Complexity → low sensitivity for ch
Variance → high complexity → high sensitivity Chaps.

low Bias & low variance

X 80% all ML/DL high training Sensitivity → Upcoming dataset

	Bias	Variance
Underfitting	high	low
Overfitting	low	higher
Good fit	low	low

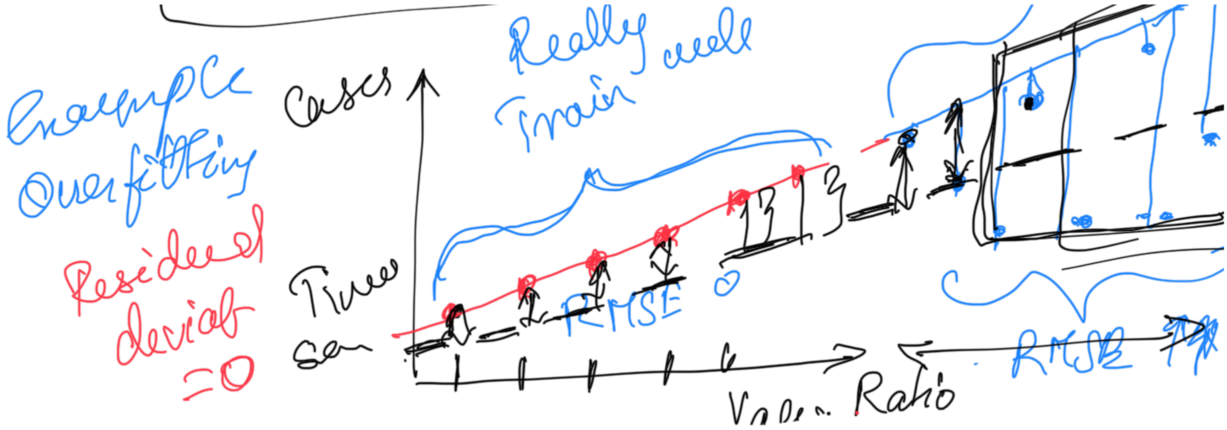
Business Case

① Logitech CTO → forecast what will be the COVID Cases 'n' months by the end of 2019

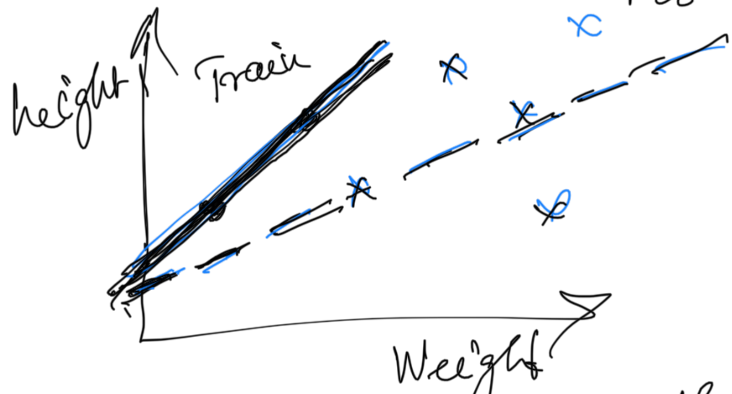
X → Vaccination Rates (Daily Basis) % Pop vaccinated, % Population

Y → No. Cases

Not Test Result Based



I am Compromising the Score on my Train to generalize the line on test.



Regularization

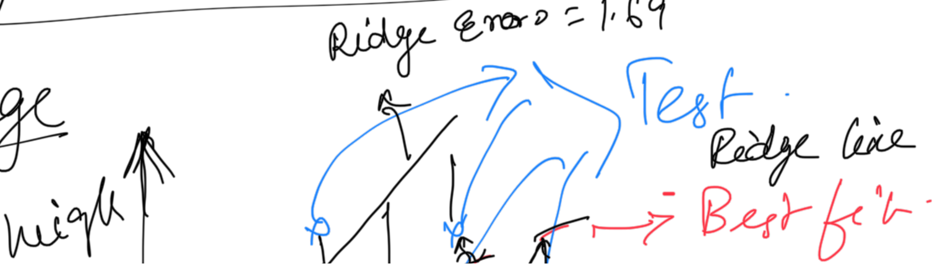
Model/Gradient finding the values of y-intercept & Slope

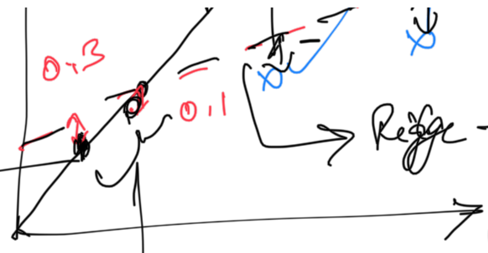
Least Square Method → SSR + Penalty

where we add penalty to SS to optimize on SS+Penalty we a regularization line

~~① Ridge~~ , ② Lasso ③ Elastic net

① Ridge





SSE

⇓

0

Ridge → 0.74

Training

height = C + m × weight

$$\text{height} = 0.4 + 1.3 \times \text{weight}$$

$$\text{SSE} = 0 + (\lambda \times \text{Slope}^2)$$

$\times 1.3^2$

Term which determine magnitude of penalty

$\lambda \rightarrow 0 \rightarrow +\infty$

SSE

$$1 \times 1.3^2 = 0 + 1.69 \Rightarrow \boxed{1.69}$$

Ridge Penalized SS

after Gradient Descent

$$\text{height} = 0.9 + 0.8 \times \text{weight}$$

$$\text{SSE} \Rightarrow 0.3^2 + 0.1^2 + (1 \times 0.8^2)$$

$$\Rightarrow 0.74$$

---

2 ways in which overfitting or underfitting can be caused

May be a factor ~~X~~

1 Model Complexity → Bias, Variance

2 Model Performance

Overfitting

Train	Test
Very good	Very poor
Very poor	Very poor
Well	Well

good fit.

(1) How will underfit

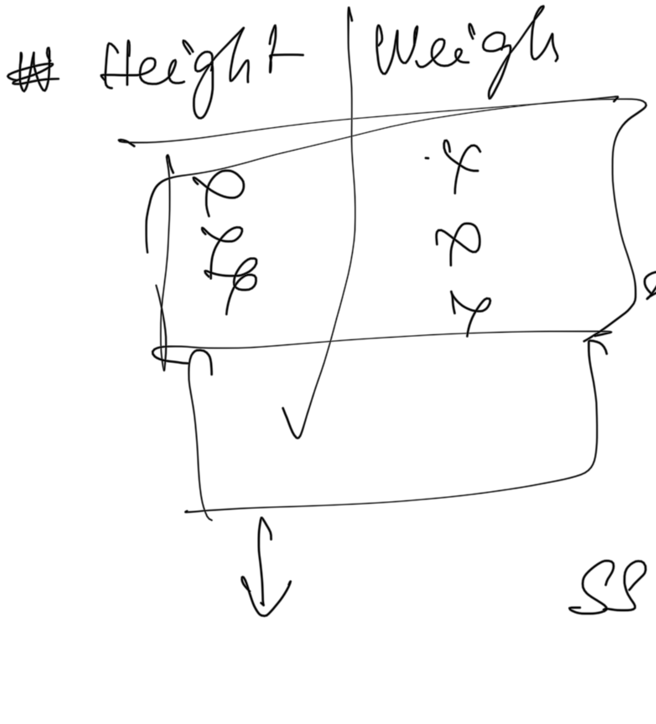
1. Add more samples
2. More feature Engg
3. Non-linear model

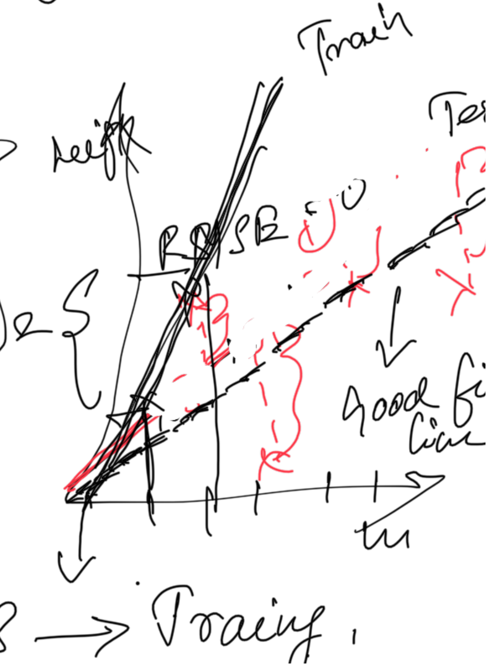
→ a bit more complexity

(2) Overfitting → Performs well on Train

Not well → Test

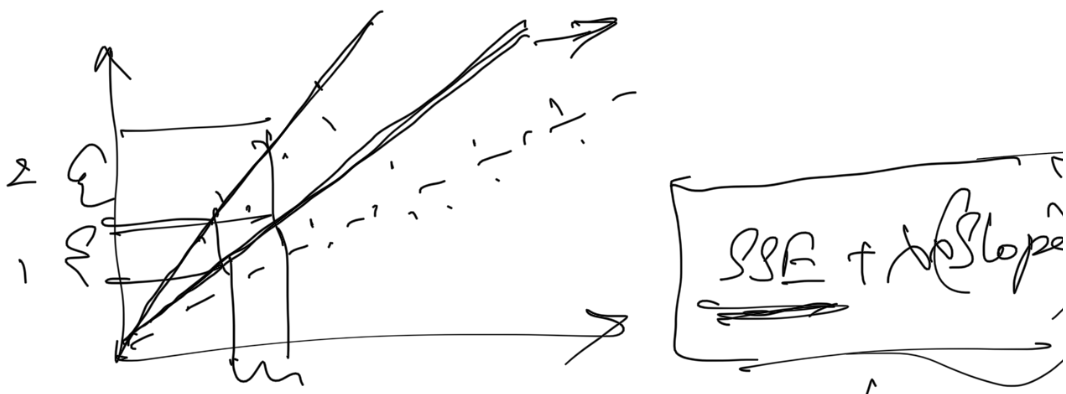
Simpler models → like LR





SS → Training.

SS + Penality.



Ridge Regres lines are less sensitive to x than least Squares L

$SSE + \lambda * (Slope^2)$

0 ⟷ +∞

Slope = 0

Linear Regsn

almost equal Mean

almost → zero.
Slope → zero.
almost.

↳ Overfitting → highly sensity to train data

→ We are Generalize → estimation

Mean of the data

Perform → 10-fold cross validation for different λ values & get the optimum λ value

tune λ

---

(2) Lasso Regularization

$$MSE + SSE + \cancel{\lambda \times Slope^2}$$

$$+ \boxed{\lambda * |Slope|}$$

Similarity

(1) Both are Generalization

(2) Both reduce the dependency of X while predicting Y

Differents

(1) Ridge → Squared penalty

Lasso → Modulus penalty

Ridge → λ → ∞

→ m ≈ 0

Lasso → λ → ∞ almost

→ m = 0 ~~Weight~~ Dietary

Weight

height = $c + m_1 x_1 + m_2 x_2 + m_3 x_3 + m_4 x_4$

0 ←

→ $\lambda(|m_1| + |m_2| + |m_3|)$

m_3 & m_4 = 0

Air Speed

Size of the house

Ridge	Lasso
① Carefully perform feature selection, columns that are significant	② a lot of columns you don't what a lot of column are there for → Lasso !!

Elastic → Both Ridge & Lasso together

300 colm, 400 colm

$\lambda_1(|m_1| + |m_2| + |m_3|) + \lambda_2(m_1^2 + m_2^2 + \dots)$

Lasso

Cross validation

(0,1) (2,2)

Ridge

(0,0) (1,5,0)

Simple → high bias & low variance.

Complex → low bias / high variance.

perfect → low bias & low variance

complexity

Bias ⇒ difference between the avg prediction of our model & correct value which.

high attention on training set

Variance → variability of the model prediction (or) data pt a value which tells us spread of data