Last class (Aug 03, 2023)

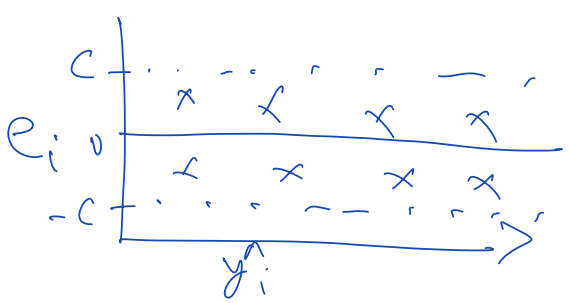
1) Recap
2) Adjusted R^2
3) Intro to stats Model
4) Assumptions of Linear Regression

Today's class

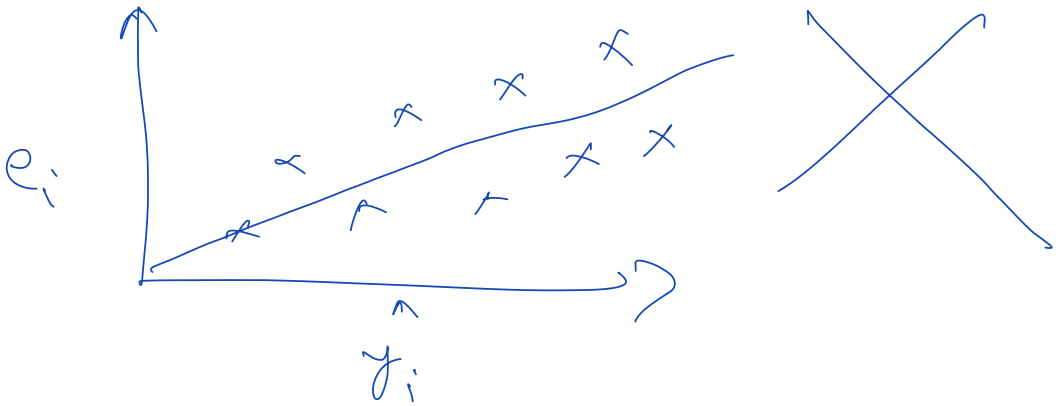
1) Quick recap
2) Overview of GD, SGD, MBGD
3) Polynomial Regression
4) Underfit vs Overfit
5) Bias Variance trade-off
6) Regularization (if time permits)

① ② ③ ④ ⑤

0.5, 0.49, 0.51, 0.505, 0.499



0 mean
Constant variance



1,000,000 Data-points

Convergence to happen very fast

↳ SGD

BGD → [m] → w. update

SGD → 1 → w. update

SGD → m samples → m weight updates

MBGD → m samples → $\frac{m}{BS}$ weight updates

1 to 1,000,000 → 1st time

→ 2nd time

→ 3rd time

SGD → 3 times → convergence

BGD → 20 times → no convergence

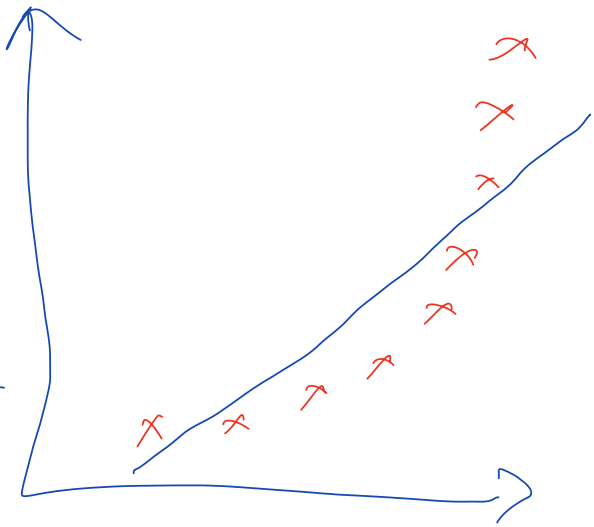
x_1 x_2 x_3

→ max variance

w_1 w_2 w_3 → highest absolute value

$x_1 \quad x_2$

$w_0 \quad w_1 \quad w_2$



$x_1 \quad x_2 \quad x_1^2 \quad x_2^2$

$w_0 \quad w_{11} \quad w_{12} \quad w_{21} \quad w_{22}$

$$\hat{y}^{(i)} = w_0 + w_{11} x_1^{(i)} + w_{12} x_2^{(i)} + w_{21} x_1^{2(i)} + w_{22} x_2^{2(i)}$$

x_1

$1/3 \leftarrow 1 \rightarrow$

$2/3 \leftarrow 2 \rightarrow$

$3/3 \leftarrow 3 \rightarrow$

x_1^2

1
4
9

$\rightarrow$

1/9
4/9
1

0.11
0.44
1

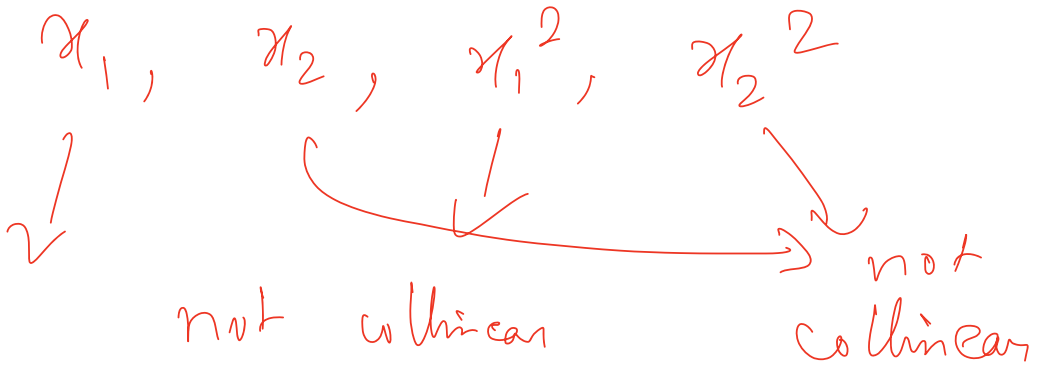
0.33
0.66
1.0

x_1 x_2 x_1^2, x_2^2, $x_1 \cdot x_2$

22

1) Feature Scaling
2) Adj R2 sure
3) MC check

$x_1 x_2 = k x_1$ ✗

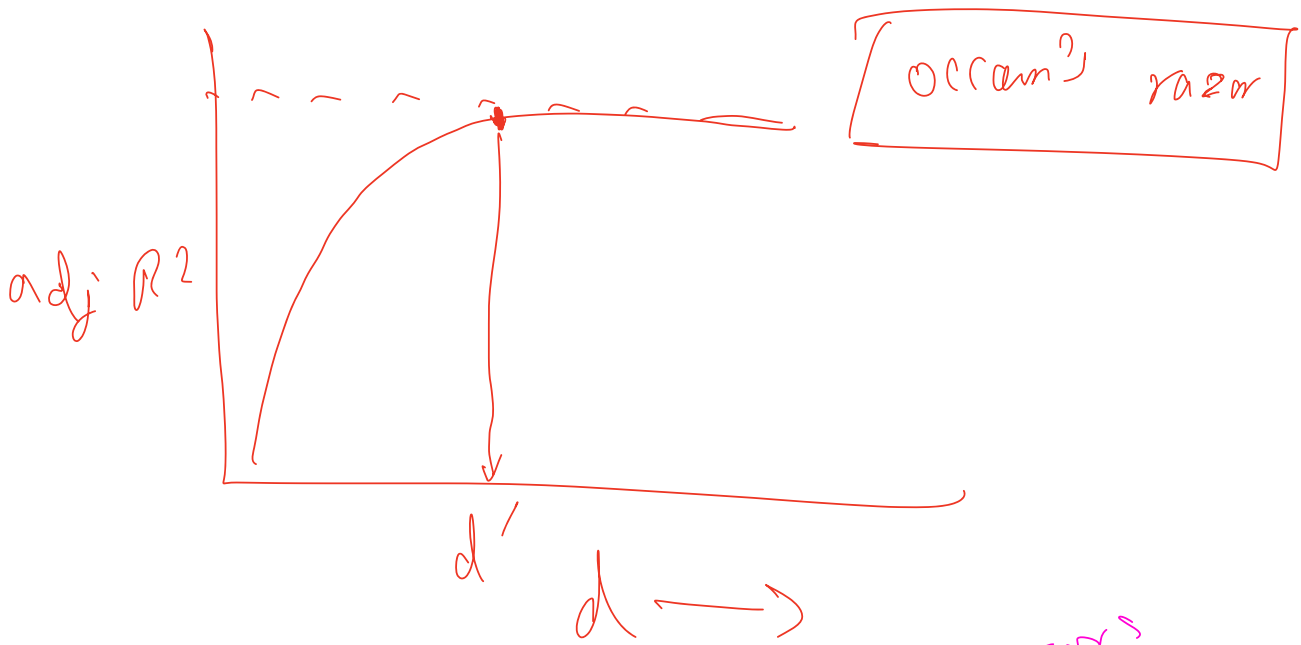


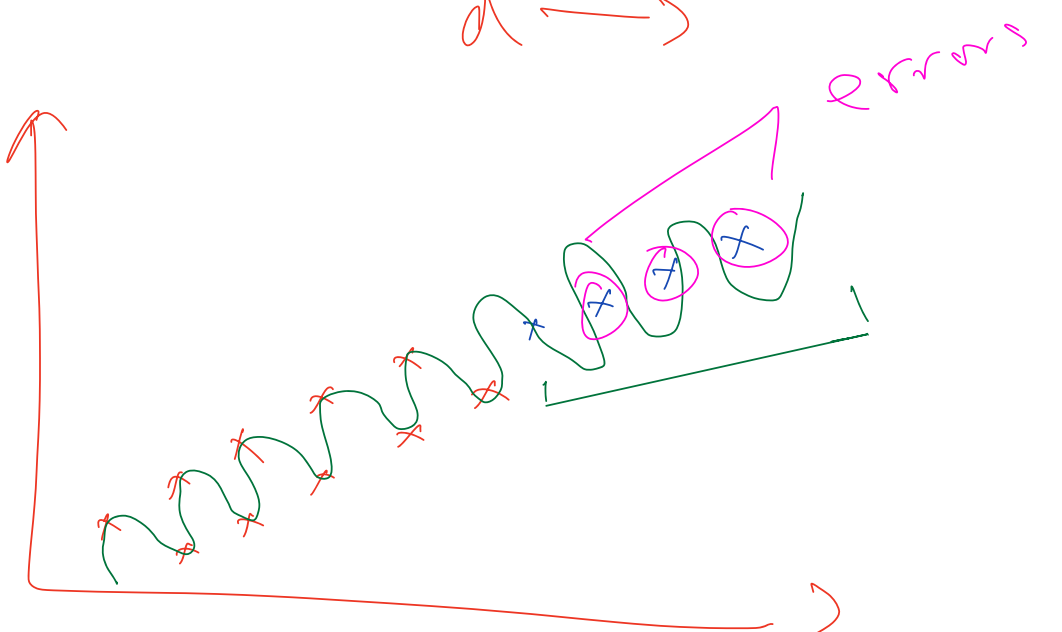
$x_1 \leftrightarrow x_2^2$

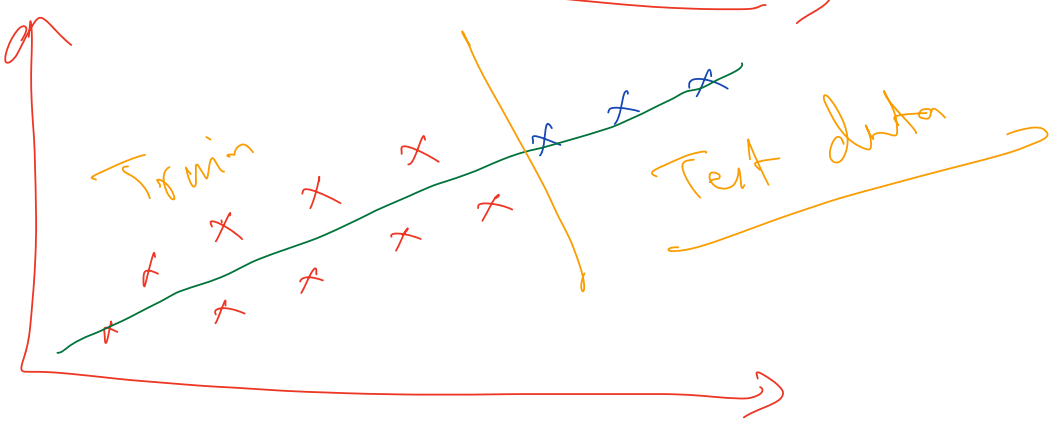
$x_1 = x_2^2$

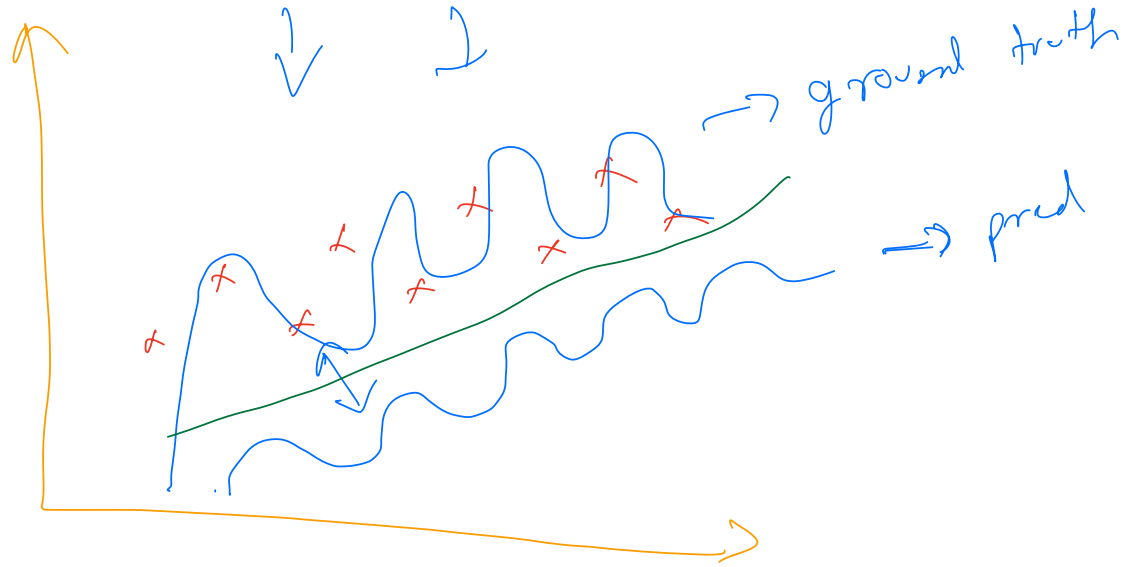
$x_2 \longleftrightarrow x_1^2$

$x_2 = x_1^2$


Occam's razor
adj. R^2
d'
d


errors


Train
Test data



Bias = mean(pred) − mean(target)

variance = mean of (Squared difference between pred & ground truth)

diagram : variance = average of difference between gd & each pred

bias : diff between avg of gd & pred

variance: 1st diff, 2nd neg

bias: 1st neg, 2nd diff

$$\text{error \%} = \frac{\text{error}}{\text{gd. truth}}$$

	Train error %	Test error %	
Case I:	40	50	High Bias, Low Var
Case II:	10	50	Low Bias, High Var
Case III:	10	12	Low Bias, low Var
Case IV:	40	70	high bias, high Var

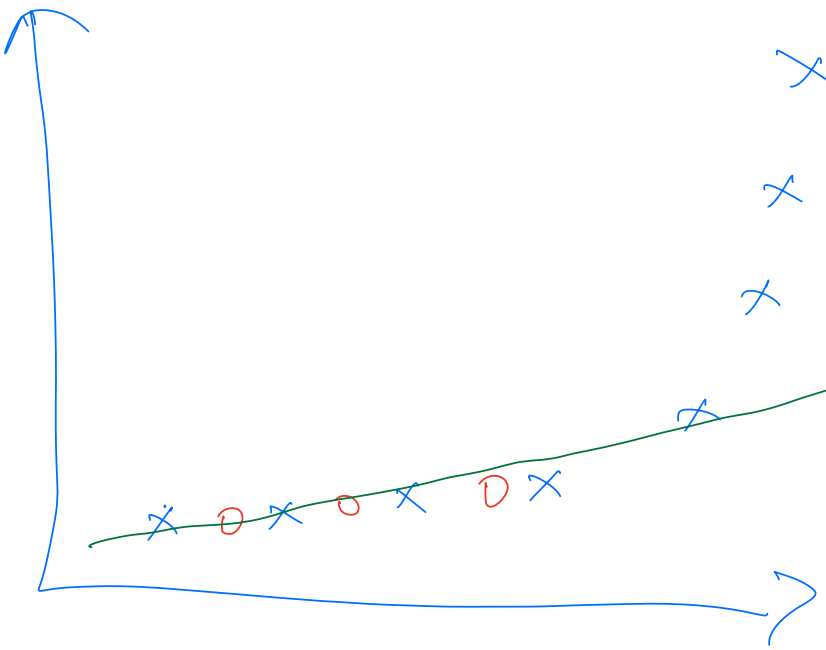
Train error → 40%

Test error → 10%

??

× → Train

O → Test



↓

degree 1

Linear

Reg

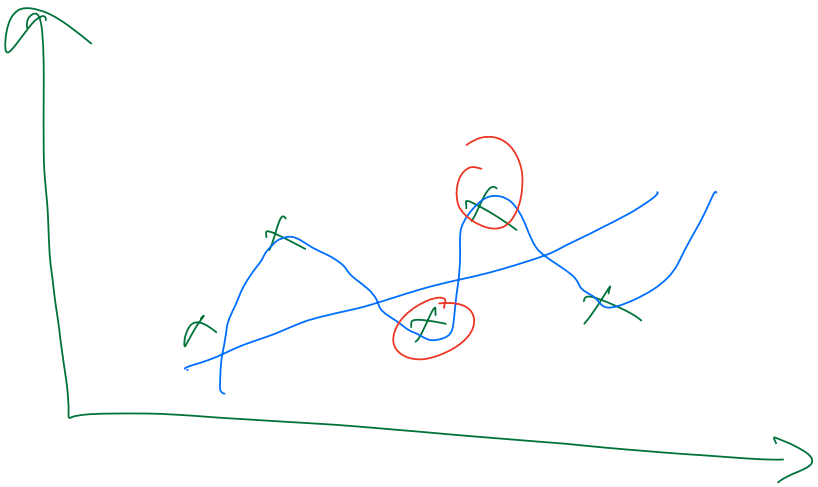
Model

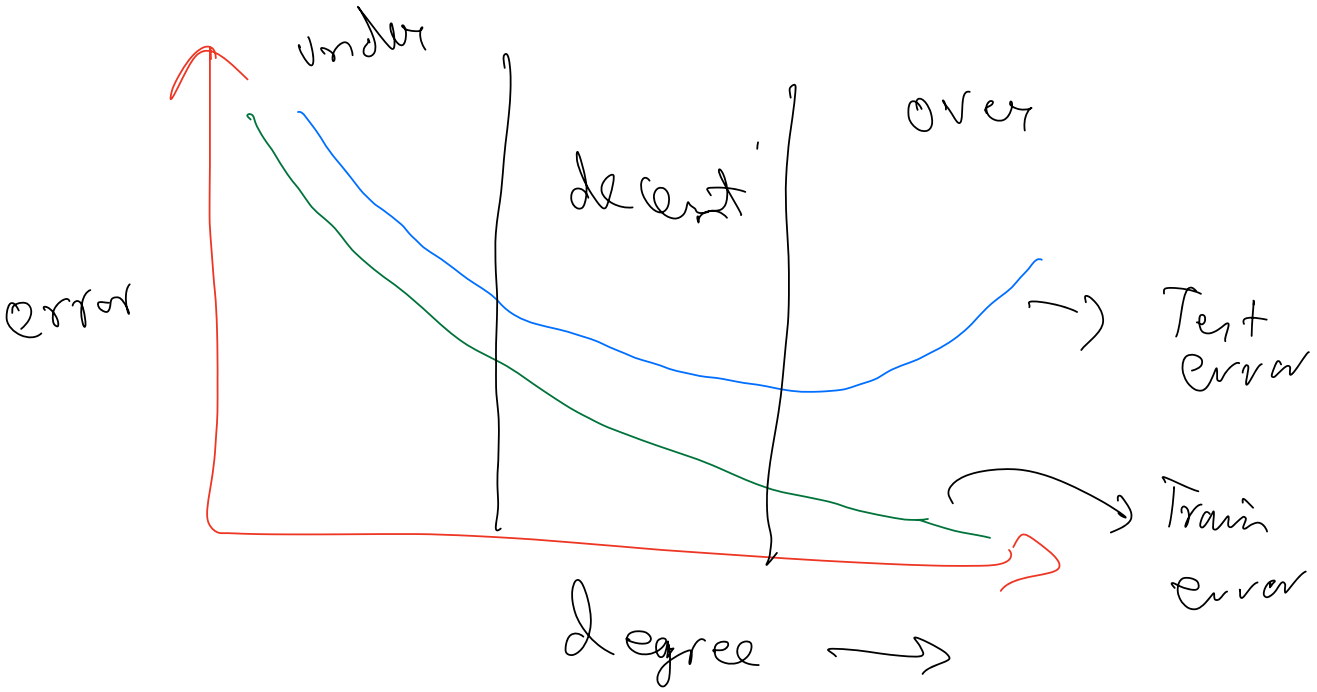
Train error → High

Test error → low

Under fitting → High Bias, Low Var

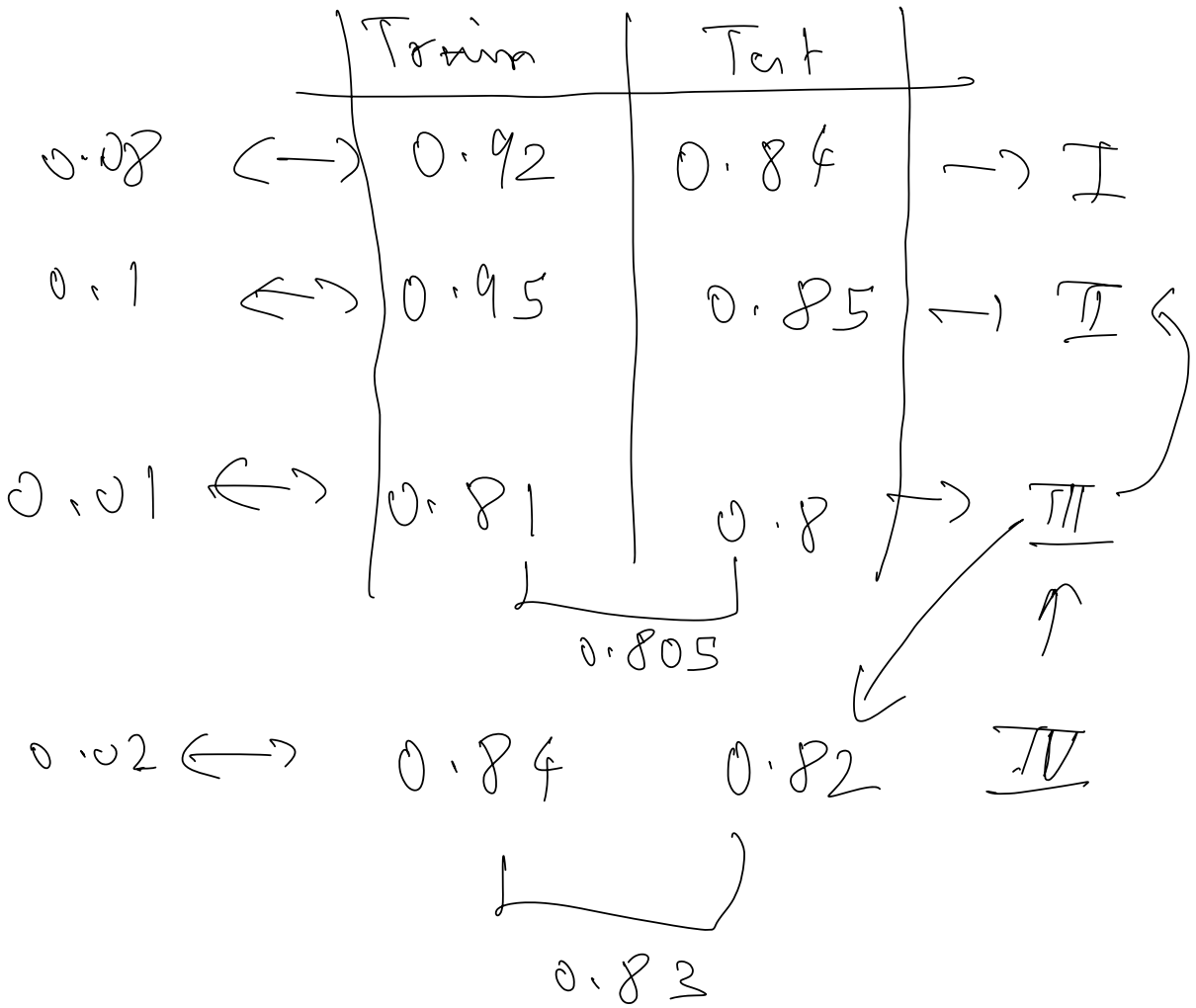
Overfitting → Low Bias, High Var





adj R2 Score :

	Train	Test
	0.9	0.8 → I
	0.95	0.75 → II
✓	0.85	0.84 → III



$0.83 > 0.805$ & III, IV are consistent

hence IV preferred

$w_0 + w_1 x + w_2 x^2 + w_3 x^3$

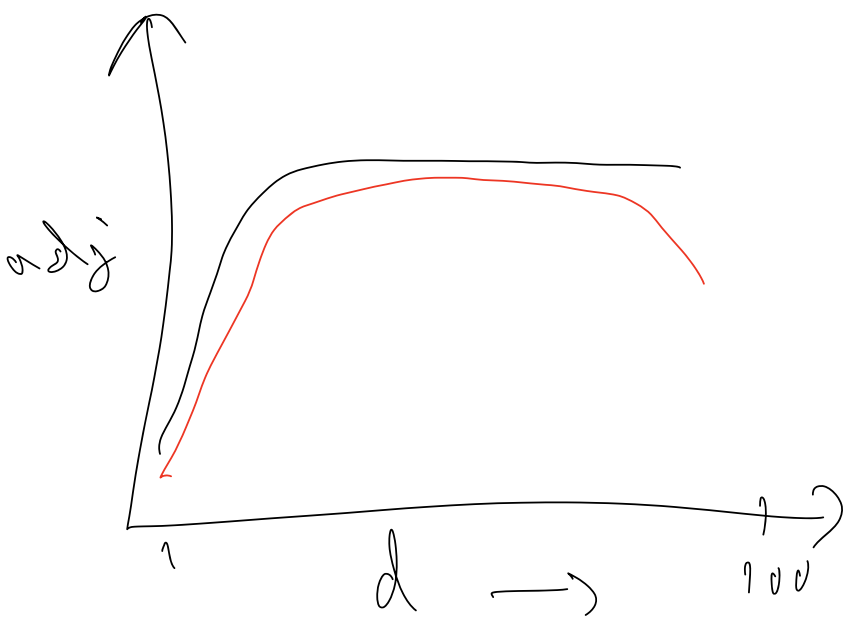
$f \ldots w_{80} x^{80} + \ldots w_{100} x^{100}$

unimp

imp $w_{81} \rightarrow w_{100}$

$\rightarrow 0.001$

$\rightarrow 0$



Steps of 1

1 → 2

2 → 4

4 → 8 →

8 → 16

$O(\log(N))$, $O(N)$ 1 to 8


B. S
Cinem