$f1 \ldots f10$

$f4 > f5 > f10 \ldots\ldots$

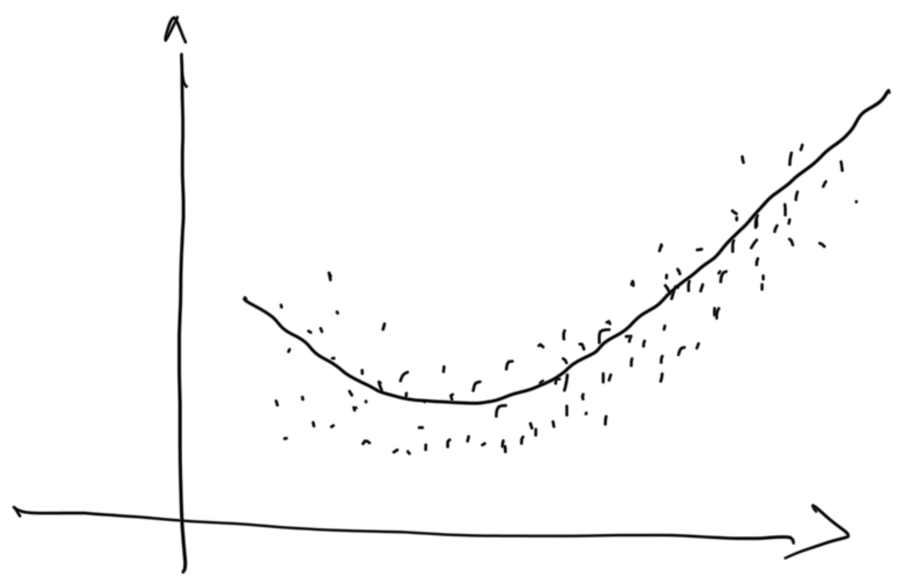
		Adj Rsquared	
SLR(f4)	—	x	
MLR(f4, f5)	—	$x + 0.1$	f5 is relevant
⋮		$x + 0.9$	
MLR(f4, f5, f8)	—	$x + 0.9$ or less	f8 is irrelevant

	r-sq	adj rsq
f5	.	.
f6	.	.
f8	.	.
f10	.	.
f1	.	.
f2	.	

seizes
irrelevant

Polynomial Regression

Non Linearity

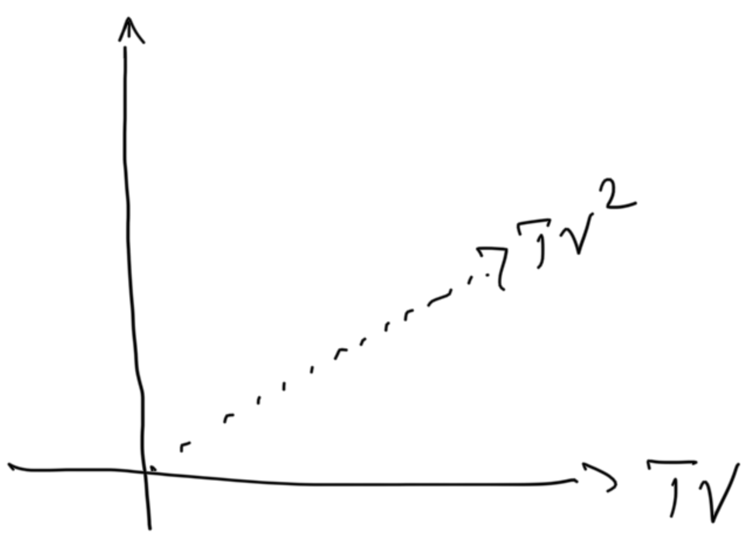


polynomial feature

TV, radio, newspaper | sales.

$$\text{sales} = \beta_0 + \beta_1 TV + \beta_2 \text{radio} + \beta_3 \text{newspaper}.$$

$$\text{sales} = \beta_0 + \beta_1 TV + \underline{\beta_1 TV^2} + \beta_2 \text{radio} + \underline{\beta_2 \text{radio}^3} + \beta_3 \text{newspaper}.$$



Now perform LR.

LR(poly features)

From $\boxed{\text{MLR to PLR}}$

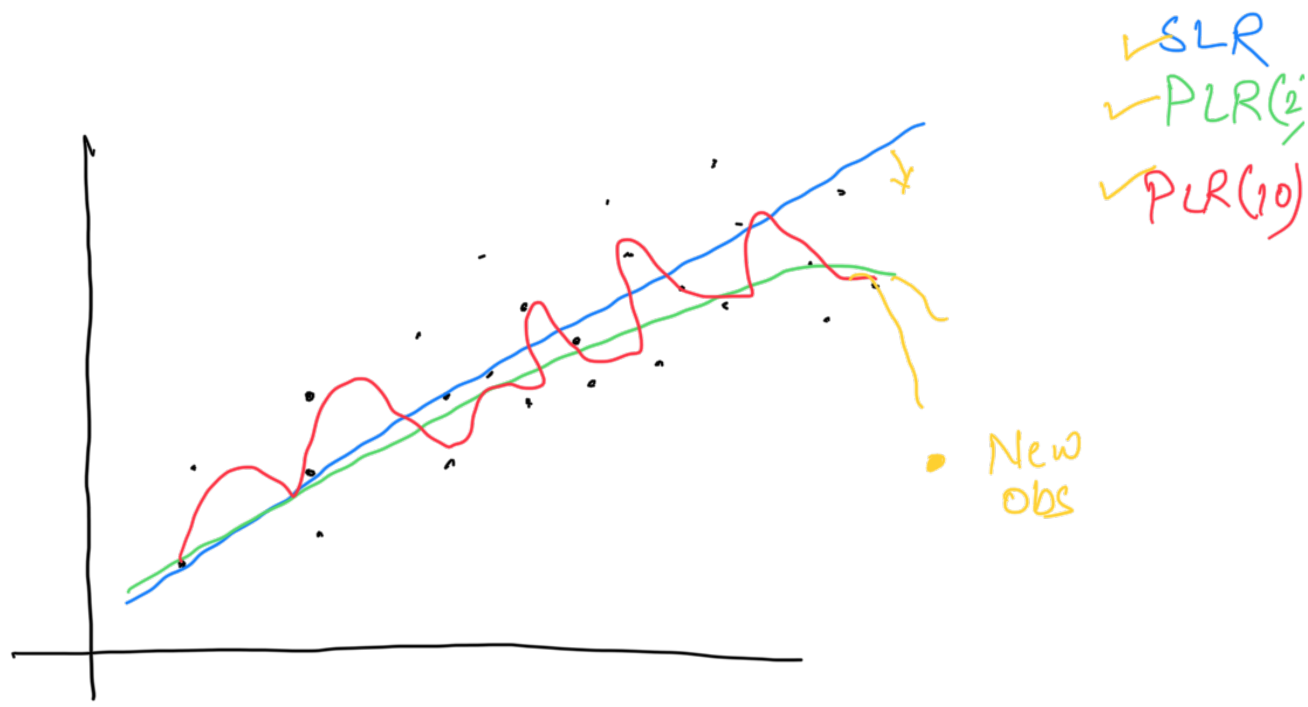
<u>change in features.</u>

order – 2 $\quad Y = \beta_0 + \beta_1 x^2 + \beta_2 x$

– 3 $\quad Y = \beta_0 + \underline{\beta_1 x^3} + \underline{\beta_2 x^2} + \underline{\beta_3 x}$

1. While exploration
2. while evaluation.

<u>Overfitting.</u>



Order of <u>sensitivity</u>

<u>PLR(10)</u> > PLR(2) > SLR

highest
→ variance
→ low bias
→ overfitting

lowest
variance
high bias
best fit

Trying to strictly fit on the train data → overfitting

Test data is different from train data.
if you strictly fit on train data then the
model starts memorizing rather than
generalizing.

Strict Teacher

2+2 = 3 → hits hard/ penalizes hard.

↓

Student will stop learning.
Memorizing. $\boxed{2+2=4}$

↓

2+3 6/4/8/

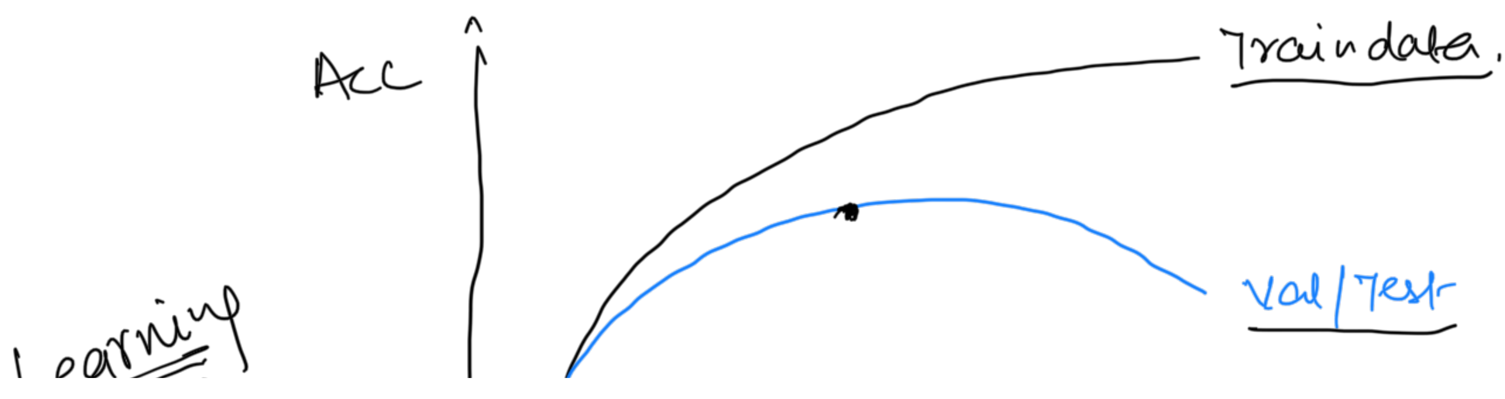
Normal teacher.

2+2=3 penalize but not hard.

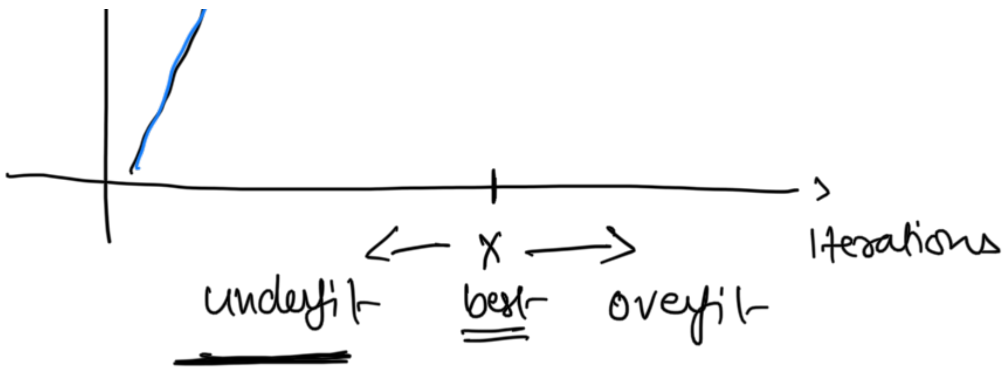
Student
Learn the logic

5

Iterations	Performance on Train data	Perform on Test data.
< 10	low	low
100 > x > 10	Moderate	Moderate
→ 500 > x > 100	Good	Good
1000 > x > 500	V. Good	Bad ↓
1500 > x > 1000	Excellent	Worst



curve



80%

keep training →

inc complexity →

Tr vs Test

① early stopping.

② Regularization

Cause :- Complexity or more training.

$y = \beta_0 + \beta_1 x$

↓

$y = \beta_0 + \beta_1 x + \beta_2 x^2$

⋮

$y = \beta_0 + \beta_1 x$

β_0, β_1

overlearn

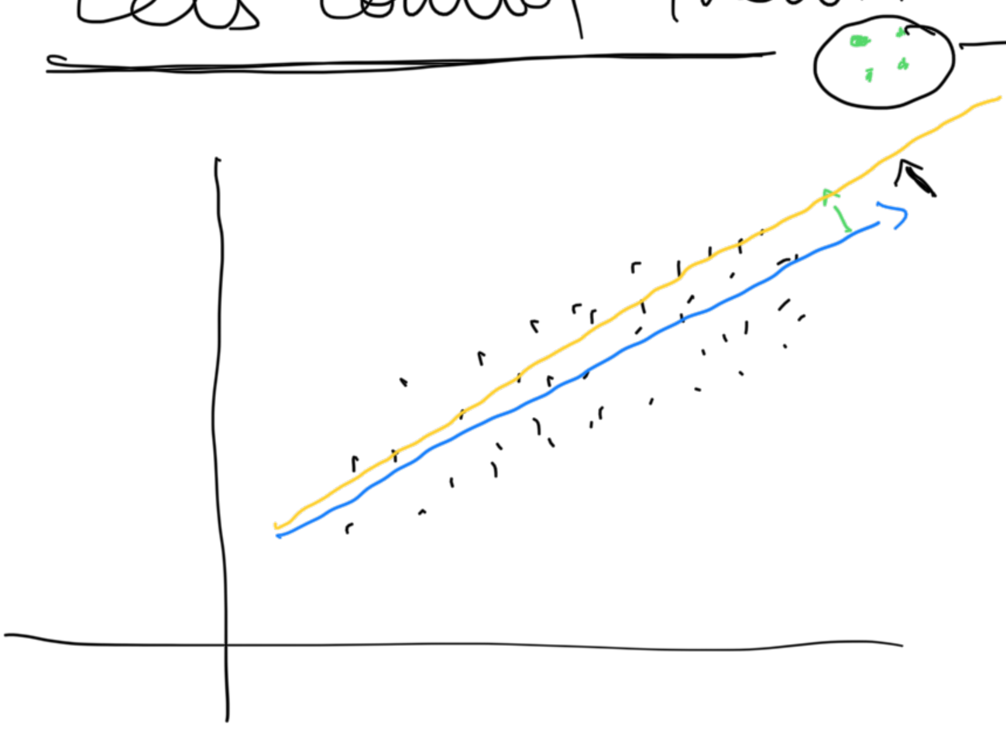
GD → No. of iterations.

$\boxed{\text{PARAMETERS}}$

↓

uncontrolled.

Lets control them.



outliers

$\beta_0 + \beta_1 x$

— Control

$\beta_2 x$

Loss function.

controlled

$$MSE = \sum (Y_i - \hat{Y}_i)^2 + \underbrace{\lambda \sum \beta^2}_{\text{regularization term}}$$

controlled

$0 < \lambda < 1$ – Reg. rate

$$\sum \beta^2 = \beta_0^2 + \beta_1^2 + \beta_2^2 + \beta_3^2 + \dots$$

$\boxed{\text{LR, MLR, PLR}}$

Regularized Linear Models.

1. Ridge regression → (multicollinearity)

$$\text{Cost function} = MSE + \lambda \sum_{i=1}^{n} \beta^2$$

L2 norm, L2 penalty

2. Lasso Regression → feature selection.

$$CF = MSE + \lambda \sum_{i=1}^{n} |\beta_i|$$

3. Elastic Net