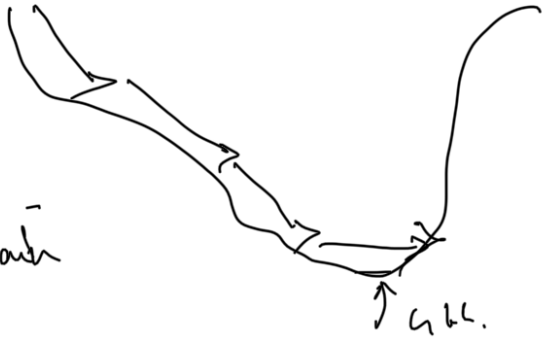
Bias variance and regularisation

① loss — quantify.

② metrics — quantify. → amount of slip → Global optim

→ Is this the final stand

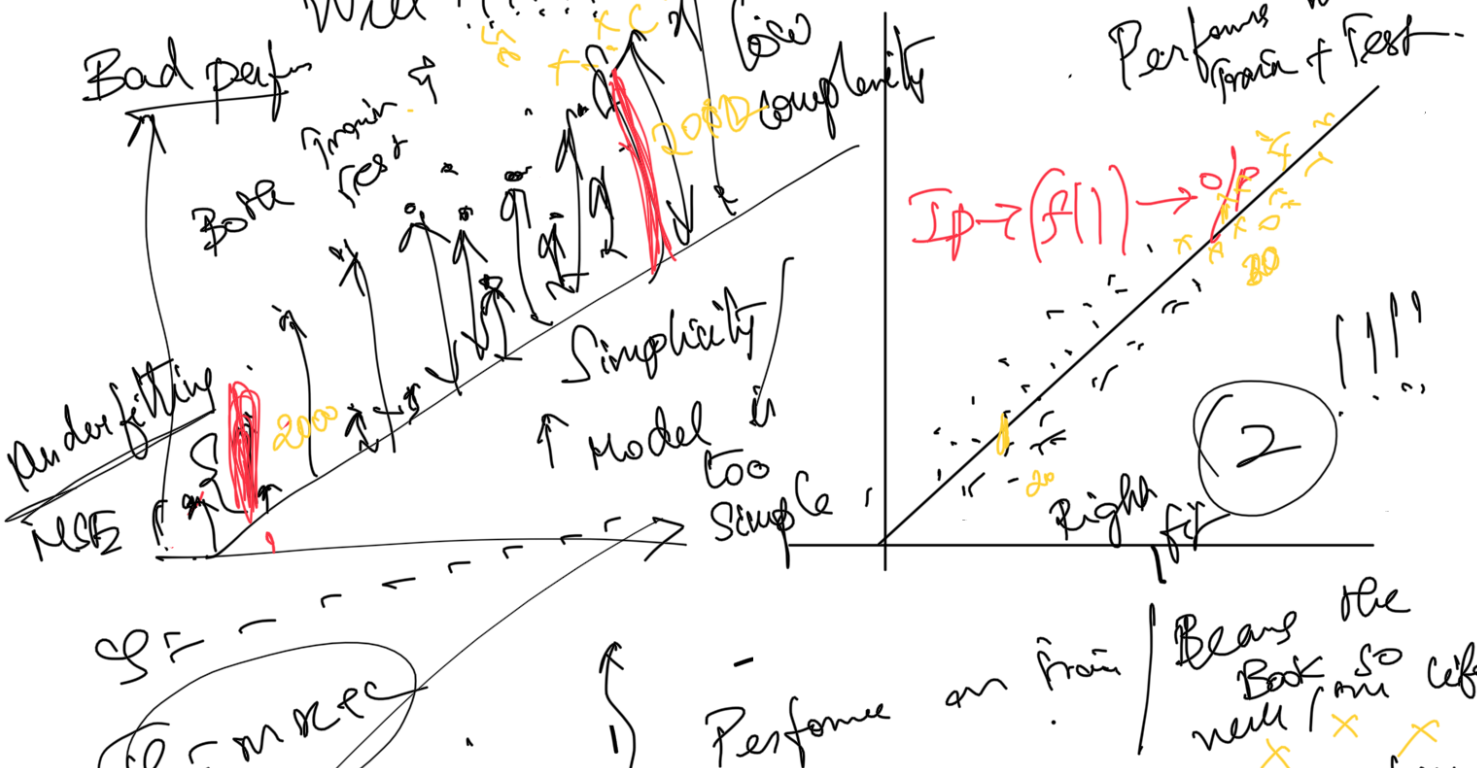


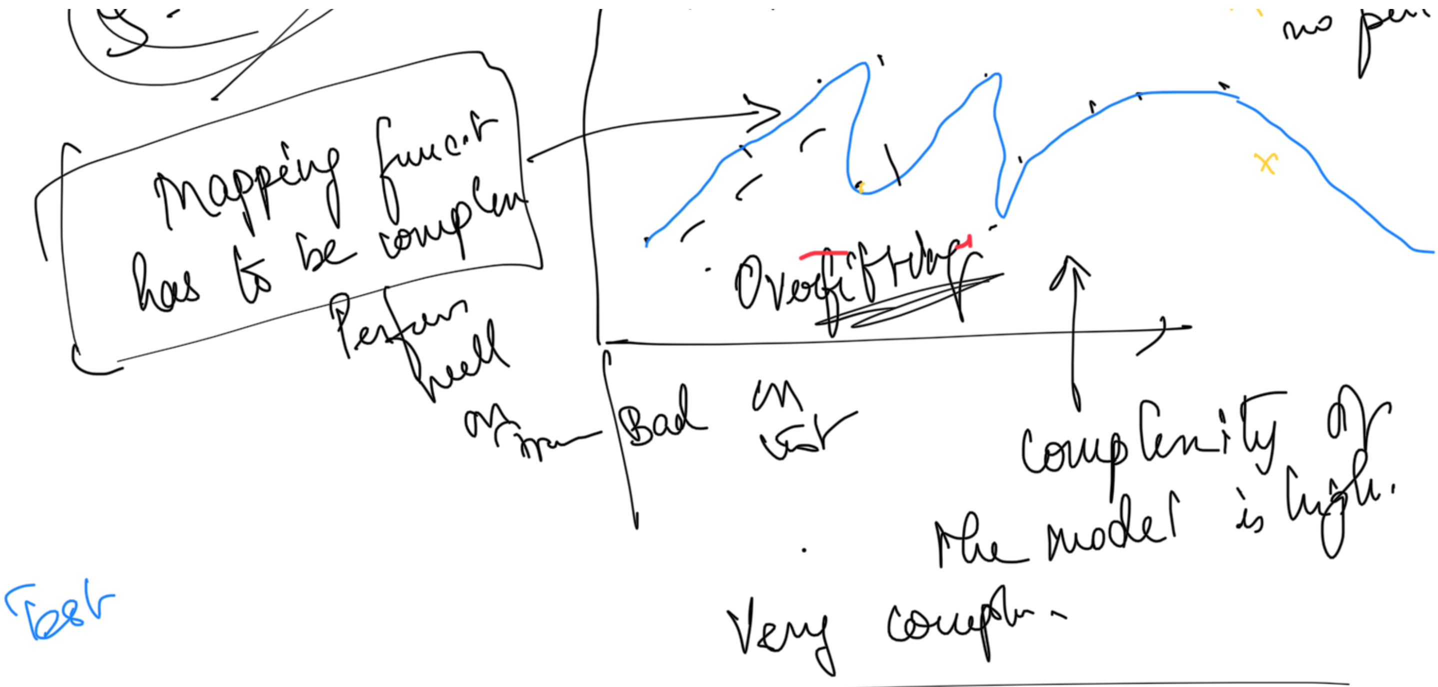
Performance on train

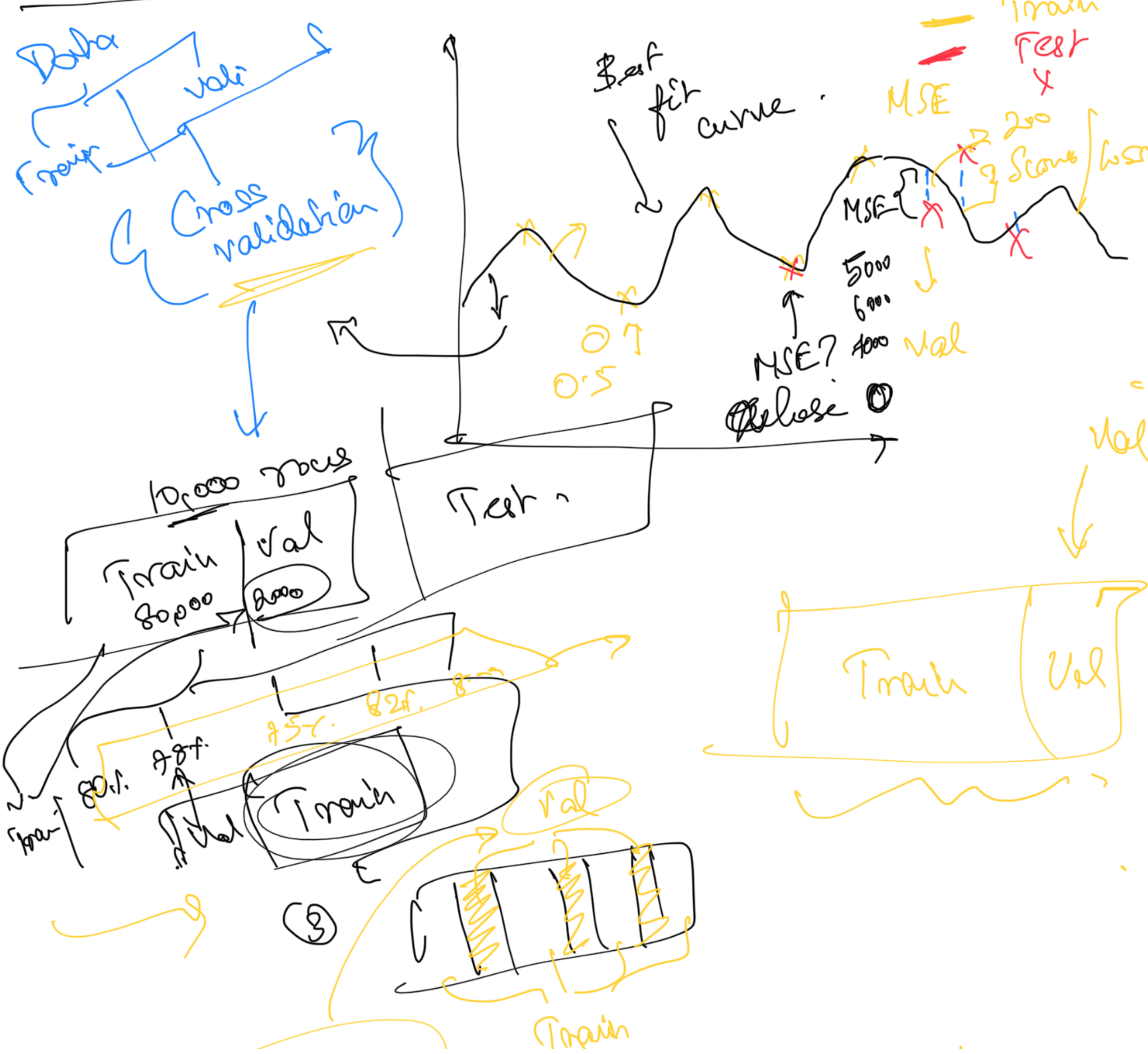
Risk Minimizer

①

Well ?????

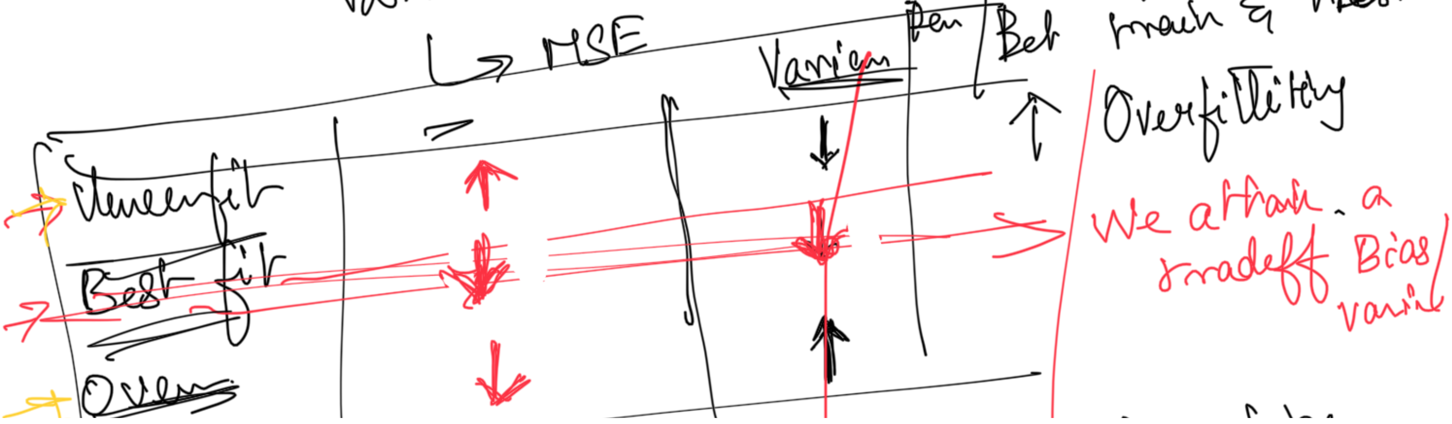


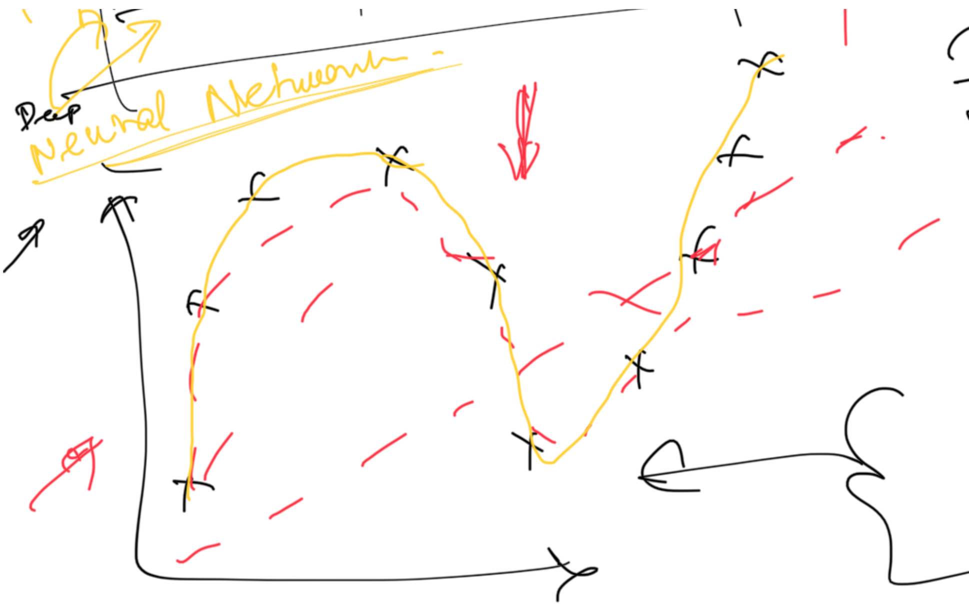

Mapping funcn has to be complex
Perform well on train
Bad on test
Overfitting
complexity of the model is high.
Very complex
Test


Data
Vali
Train
Cross validation
Best fit curve
Train
Test
MSE
200
Scores
Loss
MSE
5000
6000
4000
Val
MSE?
Release
0.7
0.5
10,000 rows
Train
8000
Val
2000
Test
80%
78%
85%
82%
Train
Val
3
Val
Train
Val
Val
Train

Variance ?? → Variance in my model

↳ MSE





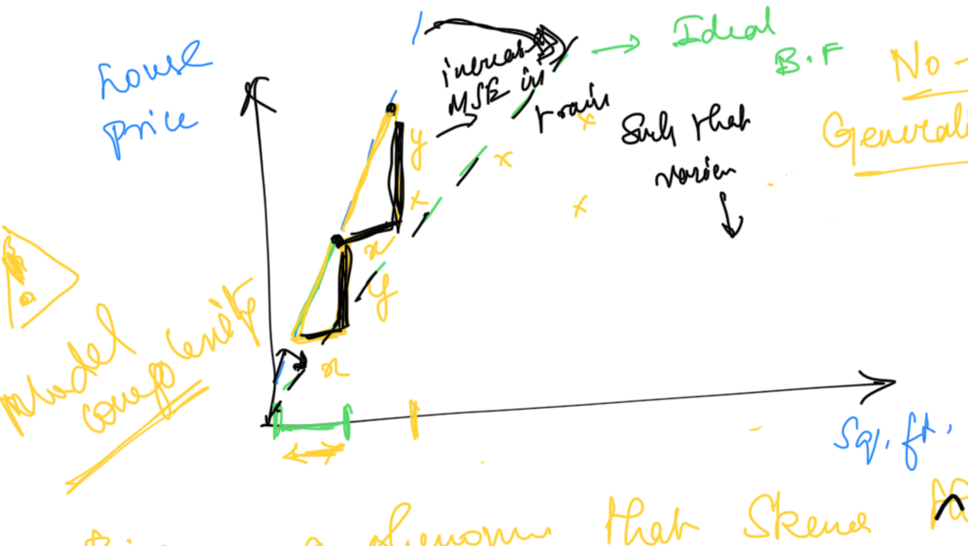
Inductive bias
↓
limitation of a learning algo
↓
Set of assumptions the the learner ...

Bias → Prejudice → occurs when a model produces o/ps that are prejudiced due to assumptions in the machine learning process
↓
Systematic erroneous

does not work in favour of training as well

(variance)

variance → Difference in the Error bet train & Test



No-Generalization!!!!

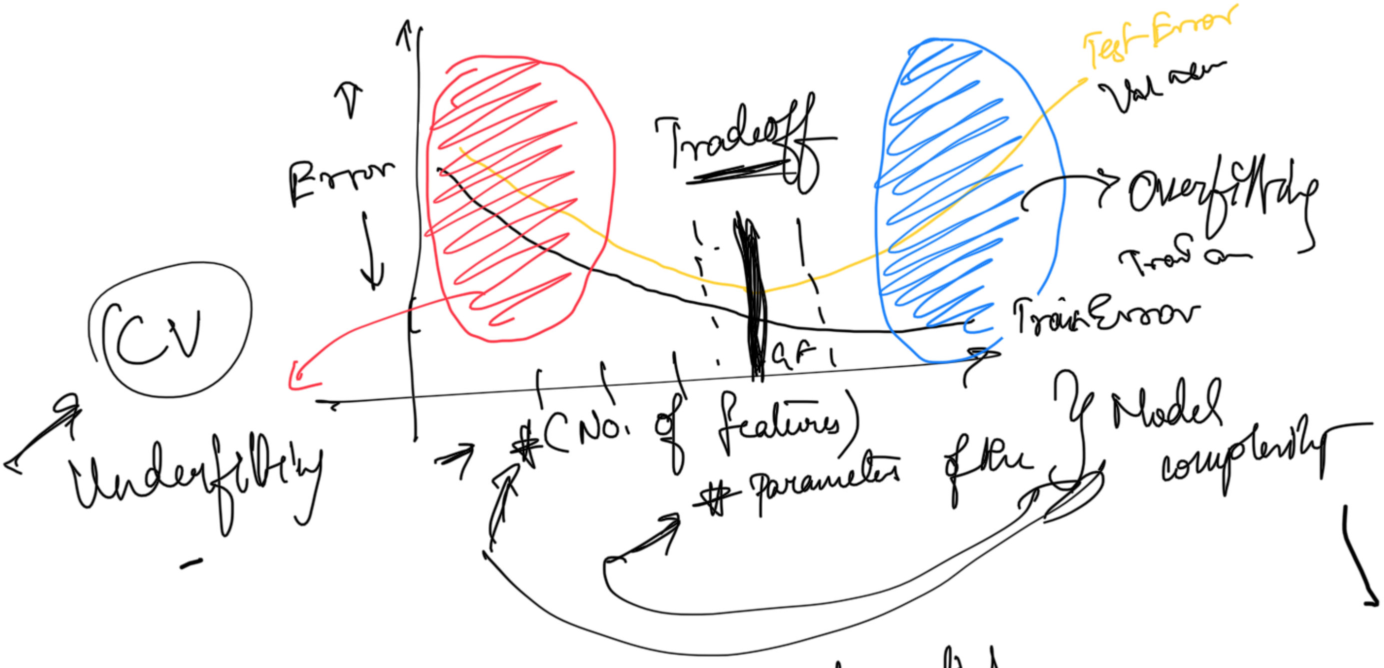
Generalize

$$\frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

Bias is a phenomenon that skews the result of an algorithm in favour or against its idea.

How do we quantify this

=> New



① Underfitting ?? Generalization can be added.

↳ adding more data / features → More complexity

↳ More complex Models → LR → SVM → En → NN

② Overfitting ?? → Regularization

My descent is too good on my train.

introduce a bias in my error term

GAN | SMOTE

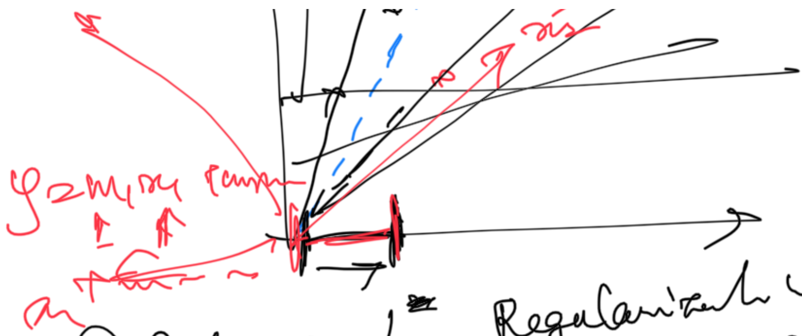
↳ MSE in train ↑

↳ Variance in MSE bet Train & Test is decr.

Gradient Descent

↳ loss => $\frac{1}{n} \sum_{i=1}^{N} (y_i - \hat{y})^2 + \_\_\_$

$= \frac{1}{n} \sum_{i=1}^{N} (y_i - \hat{y}) + \lambda [\,\ldots$



1. Lasso
2. Ridge
3. Elastic Net

(1) Ridge → L_2 Regularization,

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - (mx_i + c)\right)^2$$

Penalizing our loss term

Overfitting

$$\frac{1}{N}\sum_{i=1}^{N}\left(y_i - (mx_i + c)\right)^2 + \lambda \sum_{i=1}^{P}(mx_i + c)^2$$

For Ridge λ is parameter to exp...

↑ Training Score ↓ variance ... ↑ Train Loss, ↓ Variance (MSE) Tr & Te.

→ Slope ≠ 0 Ridge !!! asymptotically close to 0

Lasso (L_1 Regression) L_1 Regression.

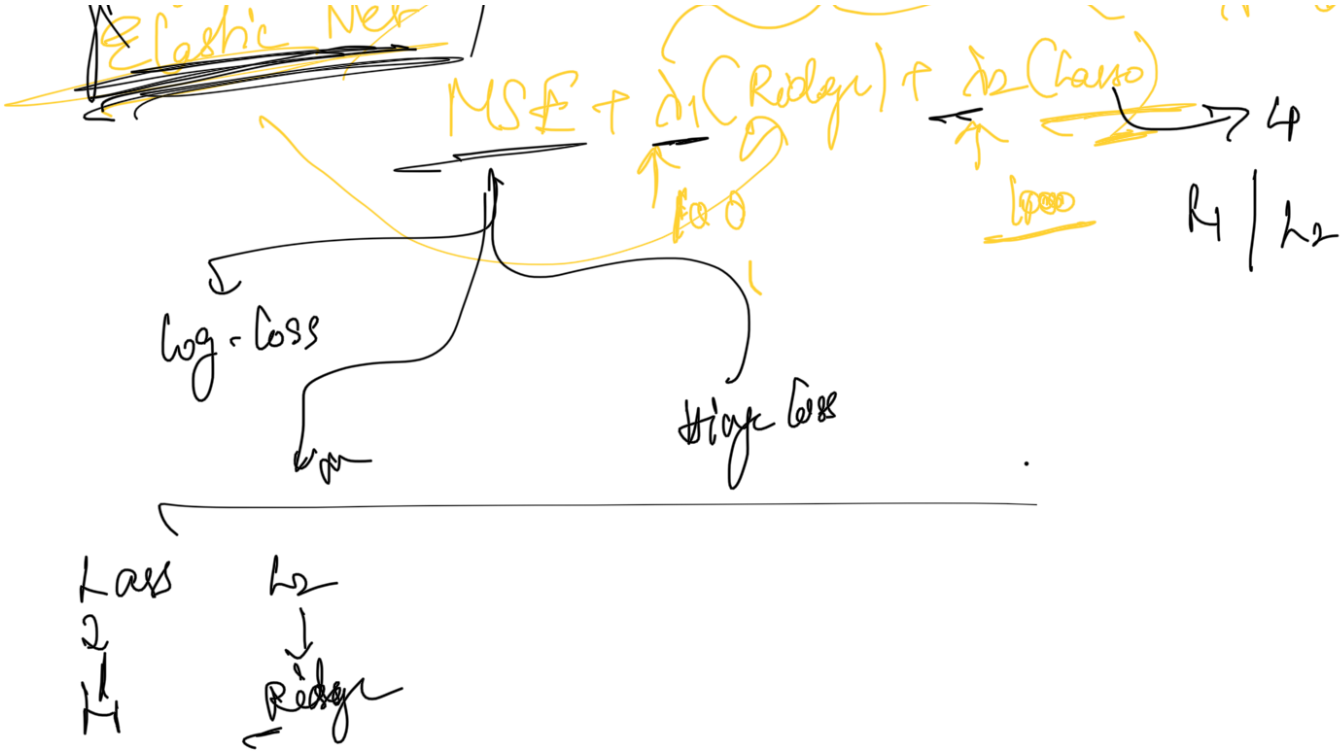
$$\frac{1}{N}\sum_{i=1}^{N}\left(y_i - (mx_i + c)\right)^2 + \lambda \sum_{i=1}^{P}|mx_i + c|$$

The big diff: ~~Slope~~ Ridge can reduce the slope close to 0 | Lasso Regularization can slowly slope to zero

Only useful columns will be retained. Useless columns will be discarded in Lasso

500 / 600

λ = ...


Elastic Net
MSE + λ1(Ridge) + λ2(Lasso)
L1 | L2
Log-Loss
Hinge Loss
Lasso
L1
L2
Ridge