

'Why regularization needed ?



model is not learning enough.

model is not flexible enough

1
High bias

High variance

→ for bagging, boosting

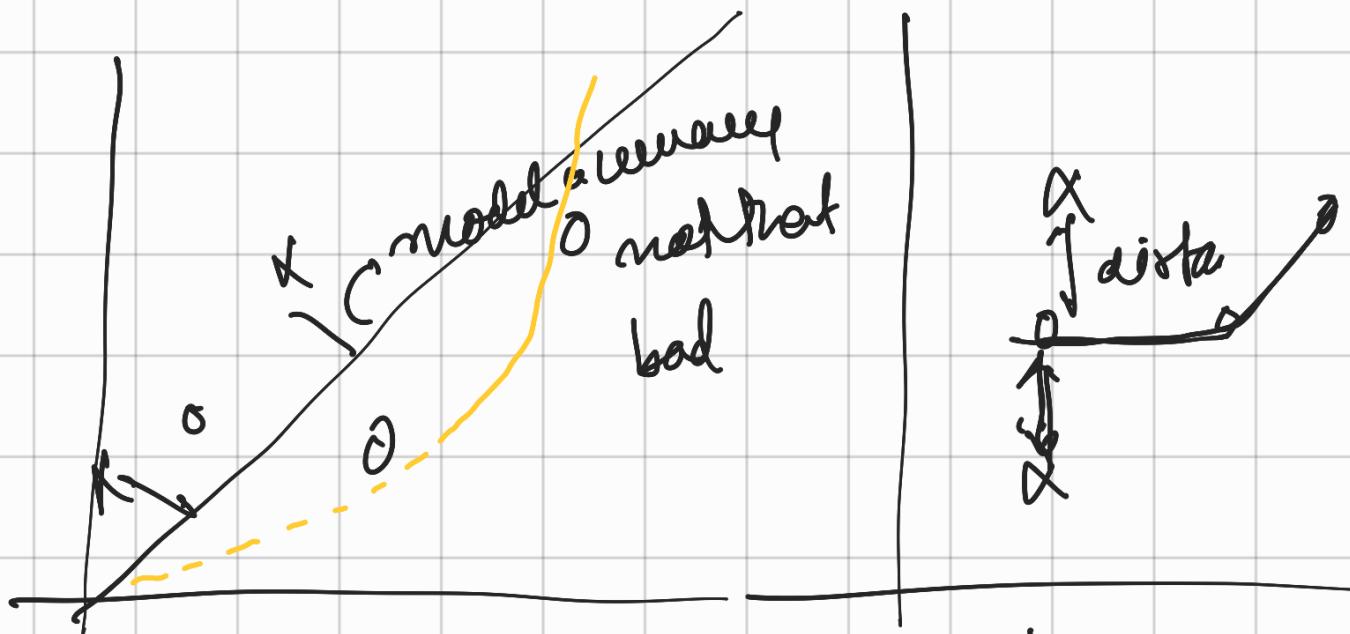
→ used in flexible

→ for dropout layer
for neural network

→ used in NN

→ used for prevent overfitting of deep learning

model



it will not give
feature as well as

but not good
on least da

l

can I have a curve with fit
properly

so there are 2 regularization function

$$\text{L2 Ridge loss} = (y - \hat{y})^2$$

~ we want to option 2

$$\hat{y}^2 = b_0 + b_1 x_1 + \text{fixed number}$$

/

$\circled{r^2(\text{slope})^2}$

so this coefficient will only make the model overfit or underfit.

this centre term

has to be low



→ the coefficient should be limited such that they

→ to keep centre function low your coefficient has to be limited such

that they should be low, hence they should not go beyond a limit.

→ Hence they are shrinkage factor add an additional term $\alpha(n)^1$

where $\hat{\alpha} \rightarrow \alpha^*$

→ for harsso $\rightarrow (\gamma - \hat{\gamma})^2 + \lambda |\beta|_{\text{l1}}$

↳ here also we shrinking the

The value of $|\beta|$ is also

with little bias

Ridge → asymptotic to 0

harsso → to equal to 0

Since harsso Regression can exclude

unless variables from equations,

it is a little better than Ridge Regression.

→ at reducing variance in the models that contains a lot of useless variables.

→ ridge regression tends to do a little better when most of the variables are useful.

Elastic Net:

- when most features are useful
 - Ridge
- when some features are useful
 - Lasso

→ When model have lot of variables

G

so how do check if we have

lasso Ridge

$$\text{sum of least square} + \underbrace{\lambda |\alpha|^2}_{1} + \underbrace{\lambda |\alpha|}_{2}$$

they have their
own lambda

When

$$\begin{cases} \lambda_2 > 0 \\ \lambda_1 = 0 \end{cases} \quad \text{lasso}$$

$$\begin{cases} \lambda_1 > 0 \\ \lambda_2 > 0 \end{cases} \quad \text{ridge}$$

$\begin{matrix} n_1 > 0 \\ n_2 > 0 \end{matrix}$ } hybrid

d

When they are correlation between
features

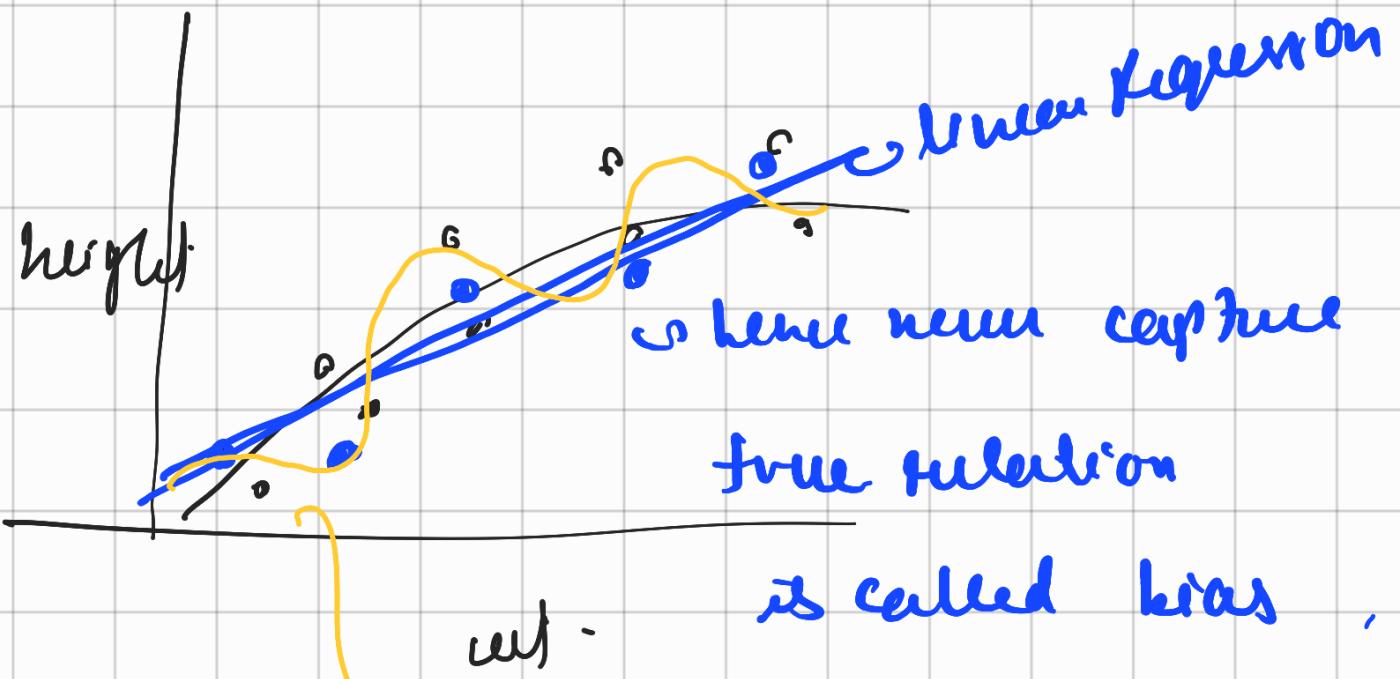
Lasso \rightarrow picks one of correlated
terms

\rightarrow shrink all the parameters

\hookleftarrow shrink the parameters correlate

with

What is Bias & Variance



→ split the training

It has very
squiggly line has super flexible

and hugs the training set along

one of , it has very little bias.

→ but did

In ML, the difference in fits



in data sets is called
variance.

→ It is hard to perform how

squiggly

→ but it doesn't fit they

well

