Laso Regression.

$L = MSE + \lambda |w|$

where $|w| = [|w_1| + |w_2| + |w_3| + \dots |w_n|$

Rigid Regression

$L = MSE + \lambda ||w||^2$

$\hookrightarrow \lambda(w_1^2 + w_2^2 + w_3^2 + w_4^2$

mean squared error / loss function

$$\boxed{\sum_{i=1}^{m} (y_i - \hat{y}_i)^2}$$

note $\boxed{\lambda = \text{alpha}}$

on Incresing the value of alpha we are minimizing $\lambda(|w|)$ this part

IF $\lambda = 0$, laso reg = normal linear reg

on low value of alpha, laso reg behave as linear reg or behave ~~under~~ overfitting

thus $\boxed{\lambda \geq 0}$

$L = \sum_{i=1}^{n} (y_i - \hat{y}_i) + \lambda |w|$

on Incresing λ 'L' must of the weight

Ridge reg ; coeff ≠ 0 (but near to 0)
but laso = 0 yes [slope hatt..]

depends on $\lambda ||w||$ or w

laso - main benifit

=> For higher ~~dimension~~ values of lambda will do Feature selection
[will zero the paticular column coeff which is not imp or y don't rely on it]

=> It also decrese dimension

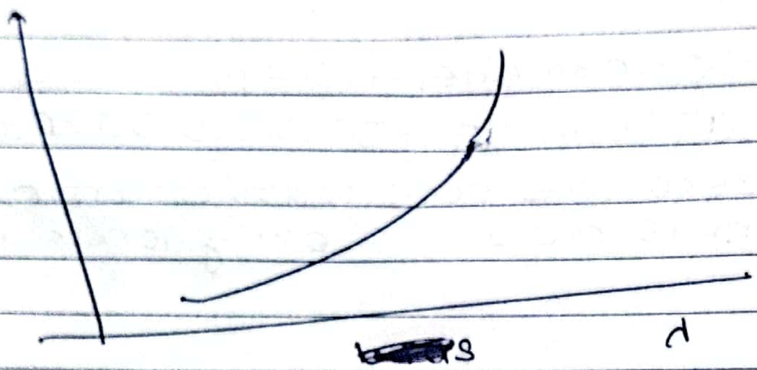
=> mostly use laso in case of rigid

x1 (x2 ↑) x3 x4 | x4 y

coeff 0 laso (say)   coef 0 laso is 0

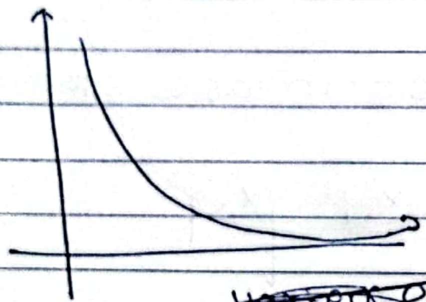
x1 ~~x2~~ x3 x4 y (Feature sel decrese dimension bias vs varience

$\lambda \uparrow$ → overfitting $\downarrow$
bias $\uparrow$
varience $\downarrow$



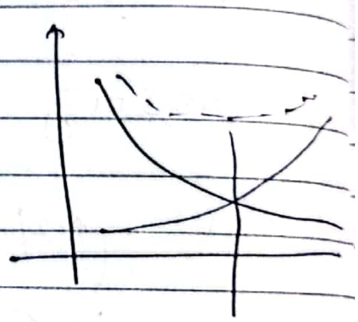
~~bia~~s

low vae of λ low bais



~~varience~~ λ

low vae of



select λ value
when bias and varien
is nearer.

Interview question

why laso creates sparsity.
(i.e $\lambda \uparrow$ $w \to 0$ why)

solu

For simple ~~m~~ Linear e

$y = mx + b$

$b = \bar{y} - m\bar{x}$

$$m = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$\bar{y} \to$ mean(y)

$\bar{x} =$ " (x)

For ridge reg

$$m = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2 + \lambda}$$

For laso — For single feature (x | y)

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda |m|$$

$$L = \sum_{i=1}^{n} (y_i - (mx_i + \bar{y} + m\bar{x}))^2 + \lambda |m|$$

(Adding 2 for only calculation)

$$L = \sum_{i=1}^{n} (y_i - mx_i - \bar{y} - m\bar{x})^2 + 2\lambda |m|$$

we know mod fun is not dif on 0,

Ujjwal barnet

so,
case → I
$m > 0$

$$L = \sum_{i=1}^{n} (y_i - mx_i - \bar{y} + m\bar{x})^2 + 2\lambda m$$

$$\cancel{\frac{dL}{dm} = \sum_{i=1}^{n}}$$

$$\boxed{\frac{dL}{dm} = 2\sum y_i - mx_i - \bar{y} + m\bar{x} \; (-x_i + \bar{x}) \quad 2\lambda = 0}$$

Rearrangment

$$\frac{dL}{dm} = -2 \sum [(y_i - \bar{y}) - m(x_i - \bar{x})] [x_i - \bar{x}] + 2\lambda = 0$$

$$= -\sum [(y_i - \bar{y})(x_i - \bar{x}) - m(x_i - \bar{x})^2] + \lambda = 0$$

$$= -\sum (y_i - \bar{y})(x_i - \bar{x}) - m \sum (x_i - \bar{x})^2 + \lambda = 0$$

or, $m \sum (x_i - \bar{x})^2 = \sum (y_i - \bar{y})(x_i - \bar{x}) - \lambda$

$$\boxed{m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x}) - \lambda}{\sum (x_i - \bar{x})^2}}$$

For $m = +ve$

Scanned with CamScanner

$mx + b$

$\therefore$

laso

For $m > 0$

$$m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x}) - \lambda}{\sum (x_i - \bar{x})^2}$$

For $m = 0$

$$0 = \sum (y_i - \bar{y})(x_i - \bar{y}) - \lambda$$

For $m = 0$

$$L = \sum_{i=1}^{n} (y_i - \bar{y})^2 + \lambda |m|$$

$$L = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$L = \sum_{i=1}^{n} \left[y_i - (\bar{y} + m(x_i - \bar{x}))\right]^2$$

$$= \sum_{i=1}^{n} (y_i - \bar{y} - m x_i + m\bar{x})^2$$

$$\frac{dL}{dm} = 2 \sum_{i=1} (y_i - \bar{y} - m x_i + m\bar{x})(-x_i + \bar{x})$$

$$\frac{dL}{dm} = -2 \sum_{i=1}^{n} (y_i - \bar{y} - m x_i + m\bar{x})(x_i - \bar{x})$$

$$= -2 \sum_{i=1}^{n} \left[(y_i - \bar{y}) - m(x_i - \bar{x})\right] (x_i - \bar{x}$$

$$\frac{dL}{dm} = -2 \sum_{i=1}^{n} (y_i - y)(x_i - \bar{x}) - m \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$0 = -2 \sum_{i=1}^{n} (y_i - y)(x_i - \bar{x}) - m \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$m \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})$$

$$\boxed{m = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

Same as linear ~~ea~~ regression

For $m < 0$

why sparsity

Three case

For $m>0$

$$m = \frac{\sum (y_i-\bar{y})(x_i-\bar{x}) - d}{\sum (x_i-\bar{x})^2} \quad - (i)$$

for $m = 0$

$$m = \frac{\sum (y_i-\bar{y})(x_i-\bar{x})}{\sum (x_i-\bar{x})^2} \quad - (ii)$$

for $m < 0$

$$m = \frac{\sum (y_i-\bar{y})\cdot(x_i-\bar{x}) + d}{\sum (x_i-\bar{x})^2} \quad - (iii)$$

From eq (i)

let $y = (y_i - \bar{y})$, and

$x = (x_i - \bar{x})$

and

let $yx = 100$, $x^2 = 50$

or,

$$m = \frac{100 - d}{50}$$

$d = 0$	$d = 2$	$d = 100$
$m = \frac{100-0}{50}$	$m = \frac{9}{5}$	$m = 0$
$m = 2$		$d > 100$
		$m = -1$

but
$d > 100$,
$m = -1$

If m is -ve then
we cannot use eq (i) right?

we need to use eq (iii)

$$\boxed{\begin{aligned} m &= \frac{yx + d}{x^2} \\ &= \frac{100 + d}{50} \\ &= \frac{100 + 150}{50} \\ m &= 5 \end{aligned}}$$

summery

$m = 2, \frac{1}{8}, 1, 0, -1, \cancel{/} 5$

starting

zero paxi m ko vae agai gardai so algorithm rokinxa yehi

(~~cross run or min m 2 b aaune~~ no'

also
$m < 0$
say $yx = -100, x^2 = 50$

$$m = \frac{-100 + d}{50}$$

$d = 0, m = -2$

$d = 50, m = -2$

$d = 100, m = 0 \rightarrow$

$d = 150, m = 1$

we should use

$$m = \frac{-100 - d}{50} \quad (\text{eq (i)})$$

$$= \frac{-100 - 150}{50}$$

$$\boxed{m = -5}$$

so '0' jan gaeo

start (-2), $m = -5$

0 tira inc hudaixa but

Scanned with CamScanner

For rigid

$$m = \frac{\Sigma (y_i - \bar{y})(x_i - \bar{x})}{\Sigma (x_i - \bar{x})^2 + d}$$

If $\Sigma (y_i - \bar{y})(x_i - \bar{x}) \neq 0$
then
on incresing any value of d
$m \neq 0$

because, laso, $d \rightarrow$ is in numerator
rigid, $d \rightarrow$ denometer

and
$m = 0$ depends on value of numerator so.

If all input colums are imp then rigid
or if you know, all colum's are not
important then.

what if you have 100 colums or more
large dataset, you don't have idea abt
colums,
then use elastic regression which
is combination of both.

$L = \Sigma (y_i - \hat{y}_i)^2 + a\|w\|^2 + b\|w\|$

note
$d = a + b$; l1-ratio $= \frac{a}{a+b}$ | $\rightarrow$ ~~...~~de

by default
$d = 1$, l1 ratio $= 0.5$ — $0.5 + b = 1$
$1 = a + b$, $0.5 = \frac{a}{a+b}$ — $\boxed{b = 0.5}$

; $0.5 = \frac{a}{1}$; $a = 0.5$

also $d_1 = \frac{a}{d}$

$a = P_1 \times d$
$b = d - a$

- also use this on multicolinarity (ग्रु?)
then,
multicolinarity :- dependend one each others,
height and weight datasets are multi colinarity ($h\uparrow, w\uparrow$) ($h\downarrow$ $w\downarrow$)

L1 ration $\rightarrow 0.9$
90% use ridge $\rightarrow$ 10% use laso