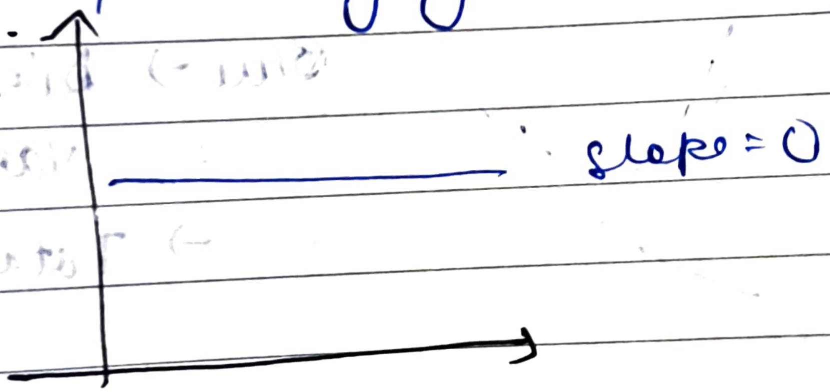
LASSO REGRESSION [L1 Regularizati

Add penalty term $\Rightarrow \lambda \|w\|$

$\lambda \left[|w_1| + |w_2| + |w_3| + \cdots |w_n| \right]$

Intutation :- Unlike Ridge, In Lasso the slope may get 0 easily



<u>Benefit</u> of slope get 0 :

In High dimentinal data there is a high chance of overfitting

If use ridge then there may be some value for some coefficients but if you use lasso, the cols that are less important there coef will be zero, ultimately you are using feature selection.

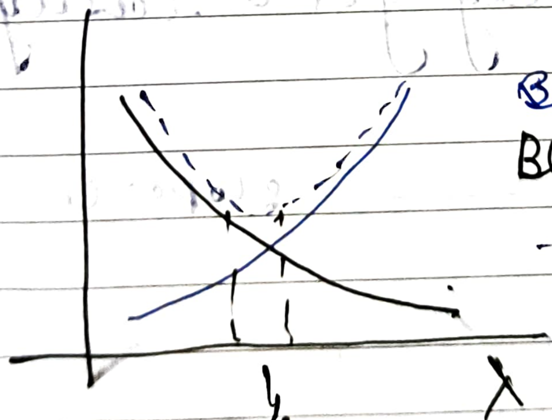
4 key points about Ridge Lasso

1. How are coefficients affected?
The coefficient will get zero, if we have a high λ value

2) Higher coefficients are affected more.

3) Features are selected at intermediate value of λ not at very high value

4) Impact on BIAS and variance.
$\lambda \uparrow$, Overfitting $\downarrow$ → BIAS $\downarrow$
→ VARIANCE $\uparrow$



Blue → bias
Black → variance
-- → Total error

In this zone you have to select value of

5) Effect of Regularization on Loss function

M T W T F S S
Page No
Date
YOUVA

curve looses its shape, and shifted at zero but never less than zero.

Q Why Lasso regression creates sparsity?

Simple linear Regression

$\rightarrow y = mx + b$

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

Ridge regression

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

Lasso Regression

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

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

$\rightarrow$ Rewrite

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

$\rightarrow$ Add $2\lambda |m|$ (calculator)

Since, it is modulus m, we can not differentiate.

$\therefore$ We can convert into cases

Case $m > 0$ Next pages

(Lasso)

case $m > 0$ $\{|m| = m\}$

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

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

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

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

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

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

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

when $m > 0$

for $m = 0$

$$L = \sum_{i=1}^{n} (y_i - \hat{y})^2 + \lambda |m| \quad > 0$$

$$L = \sum_{i=1}^{n} (y_i - \hat{y})^2, \quad \therefore \boxed{m = \frac{\sum (y_i - y_i)(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}}$$

for $m < 0$

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

so inshort in case of Lasso, λ is in numerator it can drive m to zero.