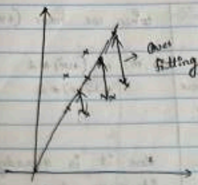
Ridge & Lasso regression :-



⇒ Ridge regression & lasso regression Gives the value which ~~given~~ is near to zero and give us value with minima error when compared to linear regression.

Over fitting - when our training dataset has no error and our test dataset has high error then it is called over fitting.

under fitting - when our training dataset has high error and our test dataset has high error then it is called under fitting.

So the main purpose of using the ridge & lasso regression is that to minimize the over-fitting that is we are reducing the high-variance to low-variance

⇒ always any data should have low bias and low-variance

Cost function :-

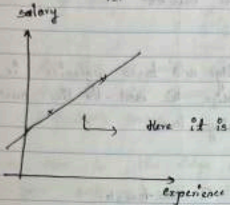
Sum of residuals

$\sum_{i=1}^{n} (y-\hat{y})^2 = 0$, ⟶ Because we fit the line with the point exactly

↓↓↓

Ridge regression

$$\sum_{i=1}^{m} (y-\hat{y})^2 + \lambda \times (slope)^2$$ Now we need to reduce this



In linear regression it is $\left[\frac{1}{2m} \sum_{i=1}^{m} (\hat{y}-y)^2\right]$

Here it is best fit line

when our experience increases our salary also increases gradually. so for this when we add our test data we will get the over-fitting problem

$$\sum_{i=1}^{m} (y-\hat{y})^2 + \lambda \times (slope)^2$$

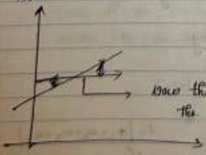
$$(0) + (1) \times (1.3)^2 = 1.69$$

as an example we took λ and slope

Because it is the best fit line

In the linear regression when we go to for the cost function we stop there

Now



Now this is not the best fit

we get some small value as cost fn

$$\Rightarrow \sum_{i=1}^{m} (\hat{y}-y)^2 + \lambda \times (slope)^2$$

$\Rightarrow$ small value $+ 1 \times (1.2)^2$

$\Rightarrow 1.44$

Now we got value as 1.44 when compared to best fit value 1.69

⇒ Now we observed that when we apply the regression with the best fit we are getting (1.69)

⇒ when we apply the regression with some value of lines which is not best fit we will get the less value when compared to best fit

The main point in the ridge and lasso regression is that we are trying to penalize the best-fit line such that the value will be coming lesser.

if we 2 slopes (or) more

it's always the same formula.

$$\Rightarrow \sum_{i=1}^{m} (y-\hat{y})^2 + \alpha \times (slope)^2$$

$$\Rightarrow \sum_{i=1}^{m} (y-\hat{y})^2 + \alpha \times (m_1^2 + m_2^2)$$

So the best fit line is the line which gets the low value

α value is always greater than 0

and it is selected using cross-validation

all are same in the Lasso regression

But the formula changes

$$= \sum_{i=1}^{m} (y-\hat{y})^2 + \alpha \times |slope|$$

$$\downarrow$$

$$\alpha \times |m_1 + m_2 + m_3 + m_4|$$

Here it is also performs feature selection such that it removes the features that are not needed and les

** In the ridge regression By penalizing the Best fit line by using formula

$$\Rightarrow \sum_{i=1}^{m} (y-\hat{y})^2 + \lambda \times (\text{slope})^2$$

So that we will change the best fit to many positions and we will choose the Best fit line accordingly when we get lower value

⇒ Here in the ridge regression we will be moving towards zero But we will not reach 0.

*** In the Lasso regression again by the same process of penalizing the Best-fit line by formula

$$\Rightarrow \sum_{i=1}^{m} (y-\hat{y})^2 + \lambda \times |\text{slope}|$$

Here it also performs the feature selection it choose the Best value and removes the unwanted values.

Here in the lasso regression we will moving towards zero and at one point it goes to zero.

Some of the features in the slope will be lossed and it reaches to zero.