MACHINE LEARNING DAY 3

15/OCT/2022

- Ridge Regression
- Lasso Regression
- Elastic Regression.

① Ridge Regression (L2 Regularization)

The main aim of Ridge Regression is to reduce **overfitting**

Note:-

Overfitting means the model will perform extremelly well for Training data but it will not perform well for Test data.

Cost Function : $\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2 + \lambda\,(\text{Slope})^2$

where λ = Hyperparameter

If $h_\theta(x) = \theta_0 + \theta_1 x$ is the equation the slope will be $(\theta_1)^2$

For multiple variables.

Cost Function : $\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2 + \lambda \sum_{i=1}^{m}(\text{Slope})^2$

Suppose the equation is $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$

Then Slope will become

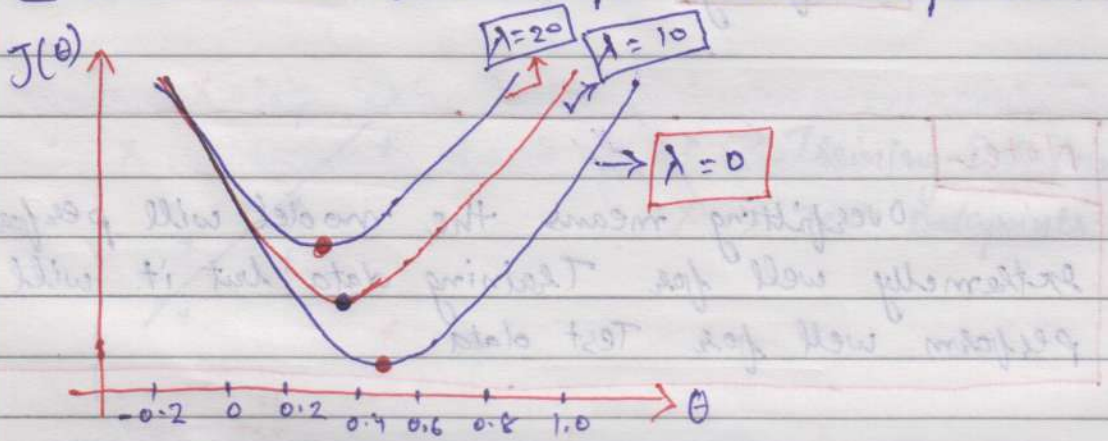
$$(\theta_1 + \theta_2 + \theta_3)^2$$

Now talk about $\lambda \sum_{i=1}^{m} (\text{Slope})^2$

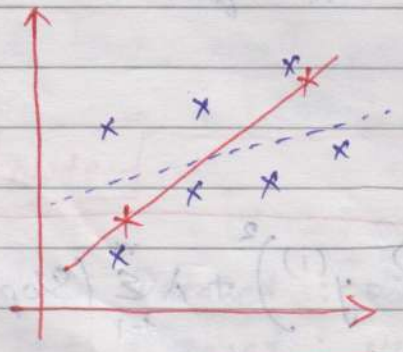
If $\lambda = 0$, cost function will be of linear regration itself.

Interview Question:-

Qs = What is the relationship between slope and λ ?



Ans:- We can say that λ is inversely proportional to slope because we can see from the above graph as the λ value increase there is a decrease in the slope.



Cost function $= 0 + \lambda(\text{Slope})^2$

$= +ve$

∴ Slope decrease as λ is +ve

Note:- In Ridge Regression, value of θ can never be Zero.

② Lasso Regression :- (L1 Norm) (L1 Regularization)

This is basically used to reduce the features.
{Feature Selection}

Formula:-

$$\text{Cost Function} : \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)} \right)^2 + \lambda \sum_{i=1}^{m} |\text{Slope}|$$

Lasso Regression uses L1 regularization technique. It is used when we have more features because it automatically performs feature selection.

Suppose:- $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$

$= \theta_0 + 0.54 x_1 + 0.23 x_2 + 0.10 x_3$

Now as value of λ increases the 3rd Feature $0.10 x_3$ decreases and at some value of λ, it will be zero and get removed therefore the model will focus only on important features.

③ Elastic Net Regression (L1 and L2 Norm)

It is the combination of both Ridge and Lasso Regression.

$$\text{Cost Func} :- \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)} \right)^2 + \boxed{\lambda_1 \sum_{i=1}^{m} (\text{Slope})^2} + \boxed{\lambda_2 \sum_{i=1}^{m} |\text{Slope}|}$$

$\lambda_1 \sum_{i=1}^{m} (\text{Slope})^2$ ⇓ Ridge Regression

$\lambda_2 \sum_{i=1}^{m} |\text{Slope}|$ ⇓ Lasso Regression