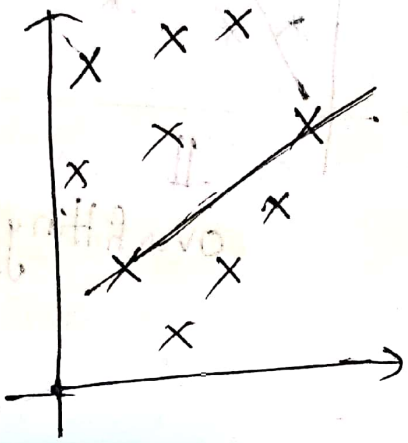
Hare Krishna.

# Ridge Regression, lasso Regression, Elasticnet Regression

## ⇒ Ridge Regression :- ($L_2$ Regularization)

↳ Reduce overfitting.



Cost function = 0

Training data low
↳ high Bias

Test data
↳ low or high Variance

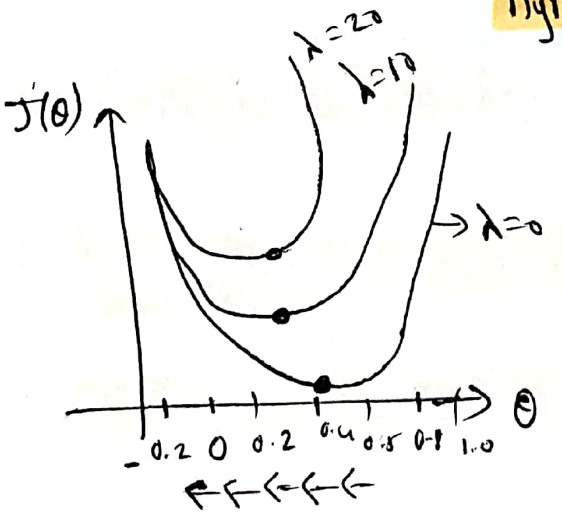
**Cost function**

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

↓

hyperparameter

{Relation b/w λ and slope}

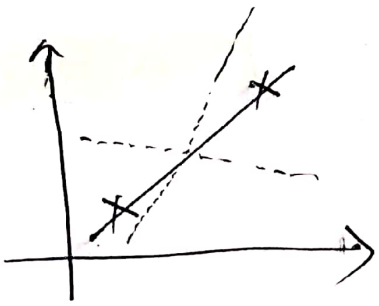



λ=0

⇓

linear regression

λ ↑ θ ↓



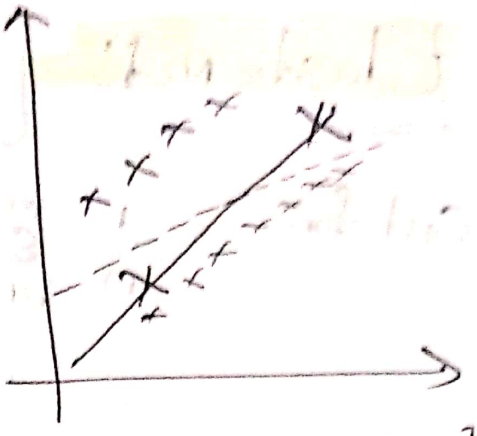
Cost fn = 0 + [+ve] = [+ve] ↓↓↓

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

$= \theta_0 + 0.45 x_1 + 0.82 x_2 + 1.5 x_3$

$= \theta_0 + 0.45 x_1 + 0.70 x_2 + \cancel{0.8}$

$\theta$ never be zero.



Cost fn = 0 + $\lambda$ (slope)$^2$

= +ve ↓↓

## Lasso Regression :-

[ $L_1$ Norm
$L_1$ Regularization ] ⇒ Reduce the fatures ⇓ feature selection

## cost function :-

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

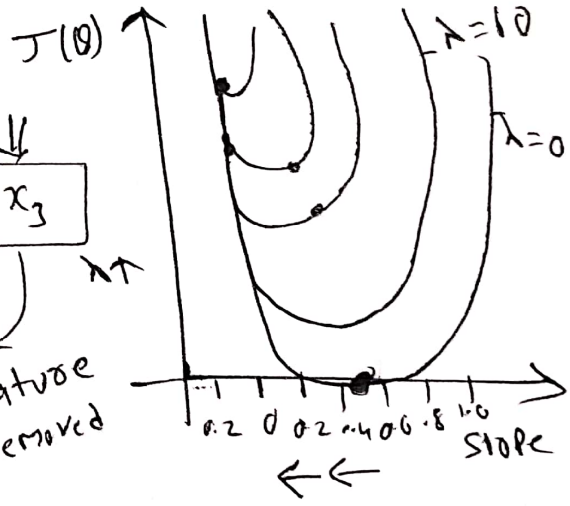
$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$

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

$y \uparrow 1$

$x_1 0.54$

$\lambda \uparrow$ ⇓ 0 ← feature removed



# Elastic Net [$L_1$ and $L_2$ Norm]

Ridge

$$\text{Cost fun} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2 + \lambda_1 \sum_{i=1}^{m} (\text{slope})^2 + \lambda_2 \sum |\text{slope}|$$

lasso