Date: 15th October, 2022

# MACHINE LEARNING (Day-3)

## Agenda:—

1) Revision of ML Day-2
2) Ridge Regression
3) Lasso Regression
4) Elastic Net Regression
5) Assumption of Linear Regression

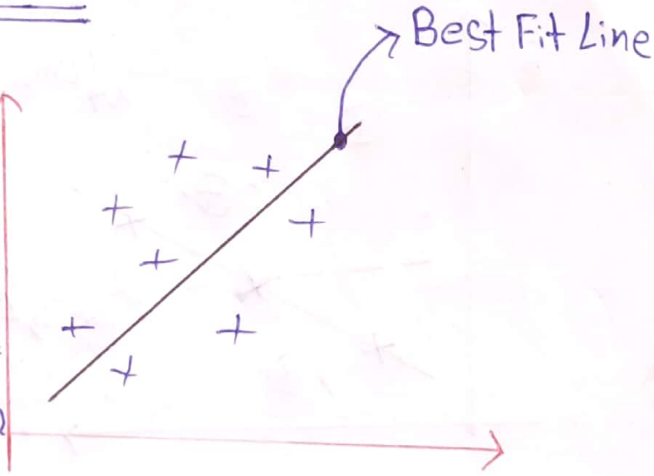
## 1) Revision of ML Day-2:—

### ★) Linear Regression:—

$h_\theta(x) = \theta_0 + \theta_1 x$

### ★ Multiple Linear Regression:—

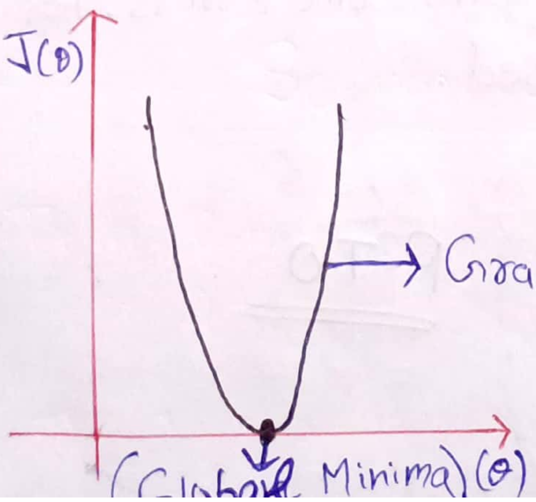
$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$




### ★) Cost Functions:— $\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2$

$\Downarrow$

MSE (Mean Squared Error)



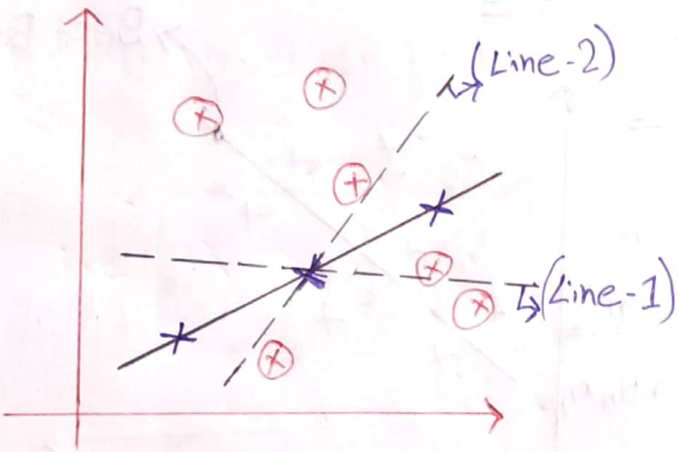

→ Gradient Descent Curve.

## 2) Ridge Regression ( $L_2$ Norm or $L_2$ Regularisation):-

⇒ The Ridge Regression is a model Tuning method that is used to analyse data that Suffers from Multicollinearity. This Method performs $L_2$- Regularisation

**Aim :-** To Reduce Overfitting.

**Example :-** Let us Consider a Sceneria where we have our train data that has overfitted best fit Line.




Here,

x → Train data (Low Bias)

⊗ → Test data (Low/high Variance)

If,

Cost Function = 0

∴ Perfectly Overfitted.

**Note :-**

⇒ Now to overcome this Overfitting Situation, we create Line-1 and Line-2 with Some Errors. For this Ridge Regression is used.

P.T.O

# *) Cost Function in Ridge Regression :-

$$\text{Cost Function} = \left(\begin{matrix}\text{Cost} \\ \text{Function}\end{matrix}\right)_{\text{Linear Regression}} + \lambda \sum_{i=1}^{m} (\text{Slope})^2$$

Here,

$\lambda$ = hypertuning Parameter
(Lambda)

Slope = Slope of individual Feature (Indepent).

$\Downarrow$

$$h\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

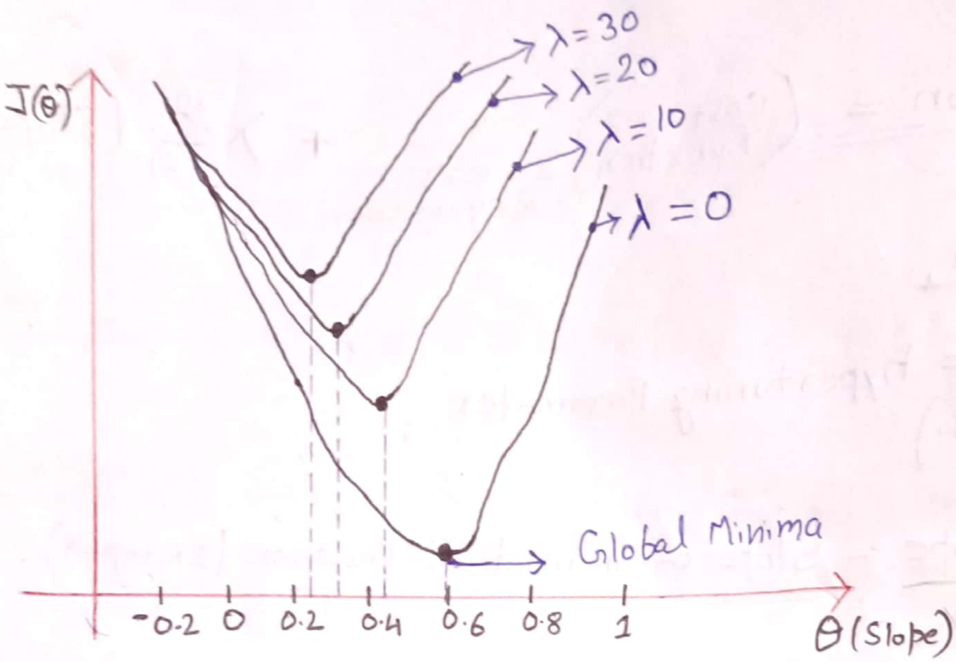
$$\boxed{\left(\begin{matrix}\text{Cost} \\ \text{Function}\end{matrix}\right)_{\text{Ridge Regression}} = \frac{1}{m} \sum_{i=1}^{m} \left(h\theta(x)^{(i)} - y^{(i)}\right)^2 + \lambda \sum_{i=1}^{m} (\text{Slope})^2}$$

$$\sum_{i=1}^{M} (\text{Slope})^2 = \theta_1^2 + \theta_2^2 + \theta_3^2 + \text{----} + \theta_n^2$$

for, $h\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \text{----} + \theta_n x_n$

P.T.O

## *) Relationship Between $\lambda$ and Slope :–




From Above graph it is evident that as $\lambda$ increases Slope decreases.

i.e $\boxed{\lambda \propto \frac{1}{\text{Slope}}}$

at worst case, $(\text{Cost function})_{\text{Linear Regression}} = 0$

$\therefore (\text{Cost Function})_{\text{Ridge Regression}} = 0 + \lambda(\text{Slope})^2$

$$= \underbrace{\lambda}_{\uparrow +ve} \underbrace{(\text{Slope})^2}_{\uparrow +ve}$$

$(\text{Cost Function})_{\text{Ridge Regression}} = +ve$ value.

$\therefore$ There will never be overfitting.

Scanned with CamScanner

**Note:-** In Ridge Regression Slope($\theta$) value will Reduce but will never reach zero.

Since, if $\theta$ reaches zero $\Rightarrow$ Feature will be Deleted.

## 3) Lasso Regression ($L_1$ Norm or $L_1$ Regularisation) :-

$\Rightarrow$ Lasso Regression is regularisation technique and it uses Shrinkages. Shrinkage is where data values are Shrunk towards central point as mean.

The Lasso Regression encourages simple, Sparse models (i.e. models with fewer features).

This modal is used when there is high Multi-collinearity or when we want to automate certain parts of model selection, like variable Selection/Parameter elimation.

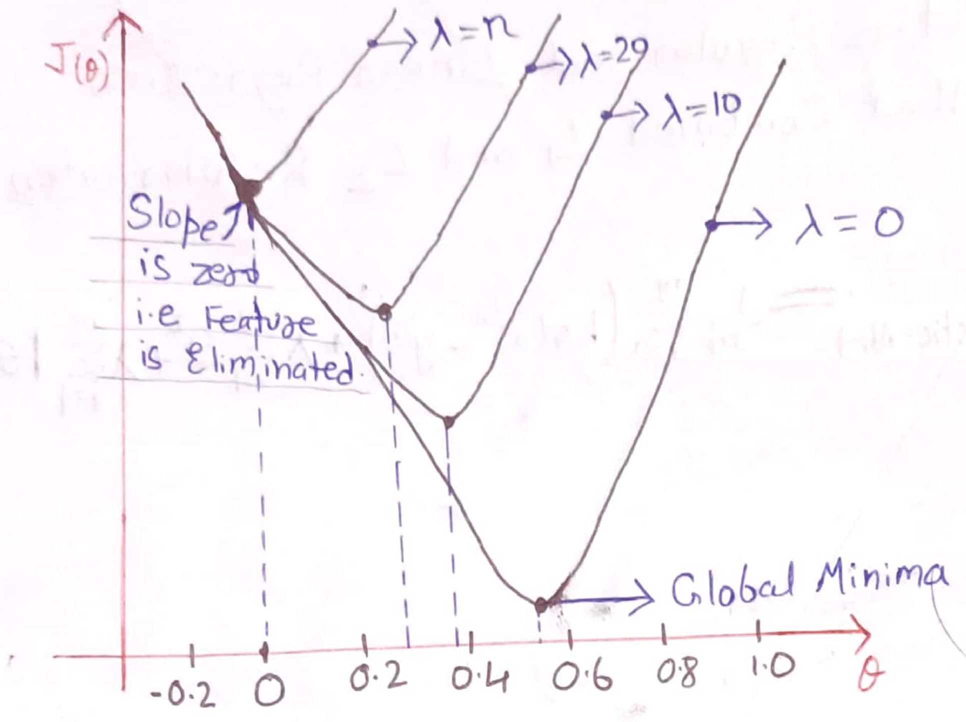
It is used when we have more features because it automatically performs Feature Selection.

**Aim:-** To Reduce Features i.e Feature Selection.

$$\left(\text{Cost Function}\right)_{Lasso} = \left(\begin{matrix}\text{Cost}\\ \text{Function}\end{matrix}\right)_{\substack{Linear\\ Regression}} + \lambda \sum_{i=1}^{m} |\text{Slope}|$$

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

# *) Relationship Between $\lambda$ and Slope :—




**Example :—** Least Correlated Features get Eliminated.

$$h\theta(x) = 22.7 + 0.54x_1 + 0.23x_2 + 0.02x_3$$

Here, $x_3$ is least correlated feature.

$\therefore$ $\lambda\uparrow\uparrow$ $\theta_3$ tends towards Zero and Finally $\theta_3$ will become Zero.

→ For Outliers we must use Lasso Regression.

P.T.O

## 4) Elastic-Net Regression :-

⇒ Elastic Net is Regularised Linear Regression technique that Combined $L_1$ and $L_2$ Regularisation.

$$\left(\text{Cost Function}\right)_{\text{Elastic-Net}} = \left(\text{Cost Function}\right)_{\text{Linear Regression}} + \text{Ridge} + \text{Lasso}$$

⇓

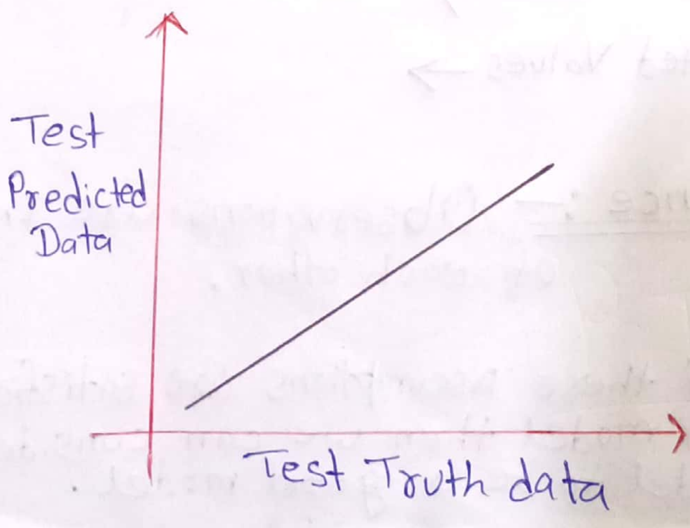
a) MSE ⓒ RMSE
b) MAE ⓓ Huber Loss.

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

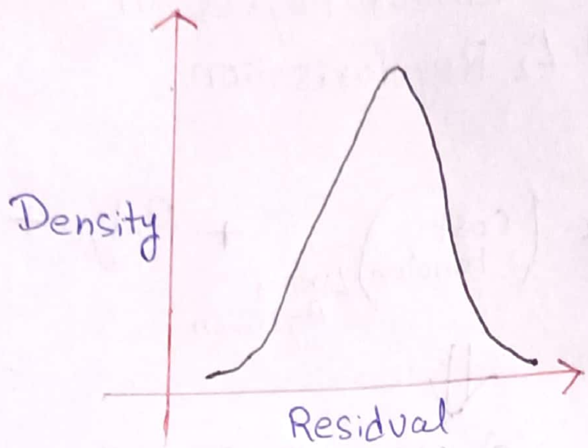
∴ In terms of handling bias, Elastic-Net is Considered better than Ridge and Lasso.

## ☆ Assumptions in Linear Regression :-

a) **Linearity :-** Test truth data($x$) and Test predicted data ($y$) must have Linear Relationship.

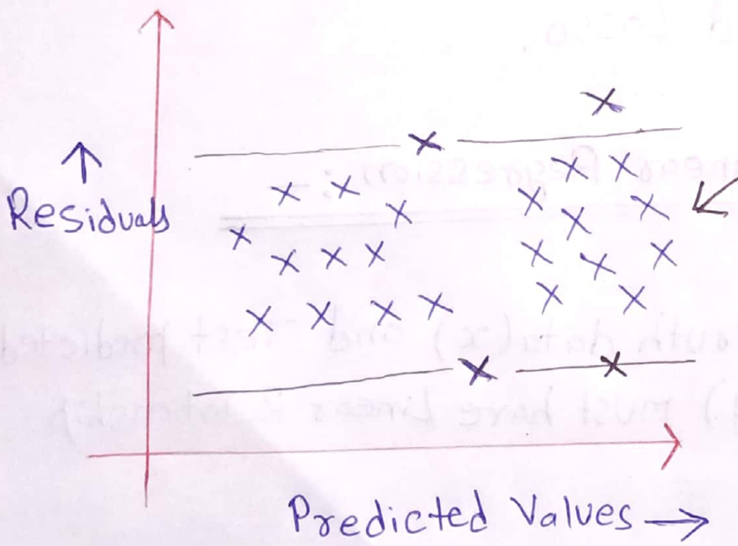


b) **Normality** :- Residual Plot Should be Normally Distributed




c) **HomoScedasticity** :- The Variance of any value of Residual is same for any value of $x$.

Scatter Plot of Predictions must have uniform Distribution with residuals.




D) **Independance** :- Observations are Independent of each other.

**Note** : If these Assumptions are satisfied by our model then we can consider our model to be a good model.