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

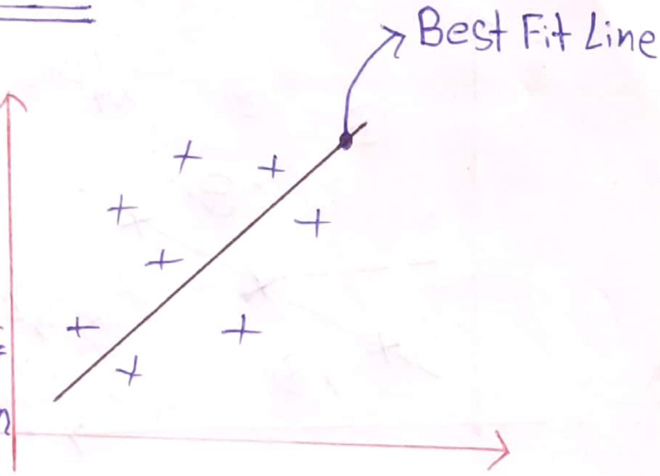
1) Revision of ML Day-2 :—

✪) Linear Regression :—

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

✪ Multiple Linear Regression:—

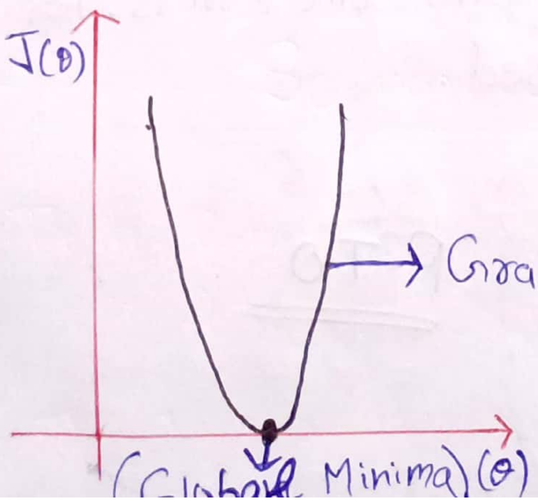
$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$



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

$\Downarrow$

MSE (Mean Squared Error)

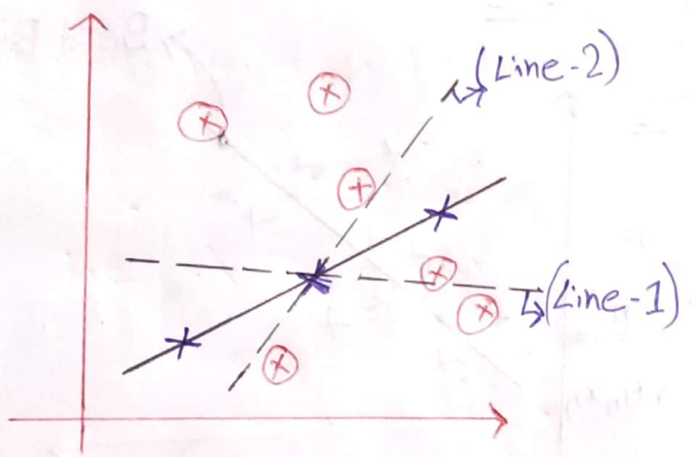


2) Ridge Regression (L_2 Norm or L_2 Regularisation):-

⇒ 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 Sceneric 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 :-

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

Here,

λ = hypertuning Parameter
(Lambda)

Slope = Slope of individual Feature (Indepent).

$\Downarrow$

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

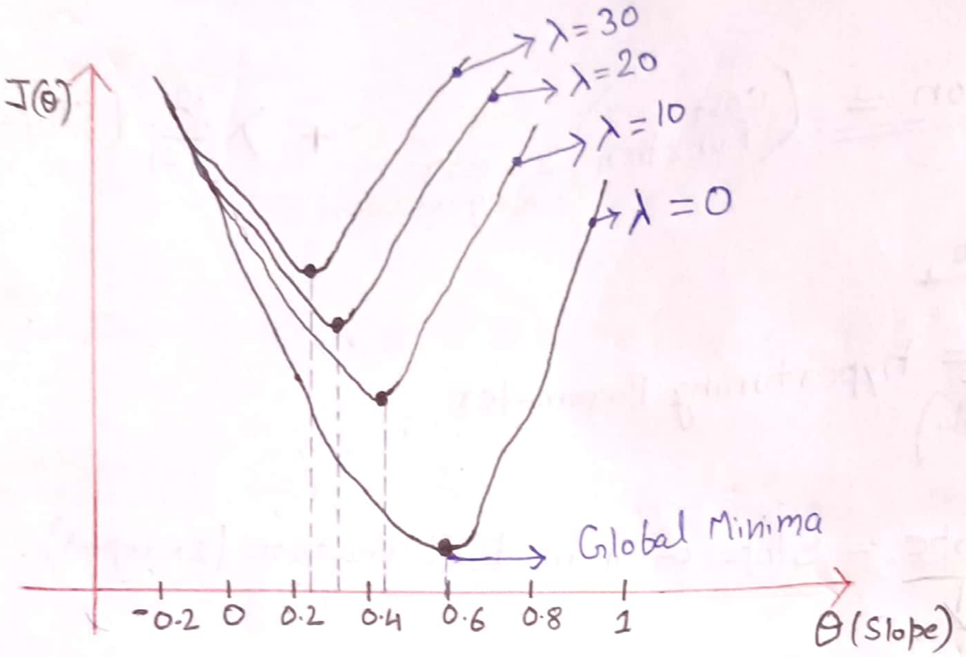
$$\boxed{\left(\text{Cost Function}\right)_{\text{Ridge Regression}} = \frac{1}{m} \sum_{i=1}^{m} \left(h\theta(x)^{(i)} - y^{(i)}\right)^2 + \lambda \sum_{i=1}^{m} (\text{Slope})^2}$$

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

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

P.T.O

*) Relationship Between λ and Slope :-



From Above graph it is evident that as λ increases Slope decreases.

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

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

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

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

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

$\therefore$ There will never be overfitting.

Note:- In Ridge Regression Slope(θ) value will Reduce but will never reach zero.

Since, if θ 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.

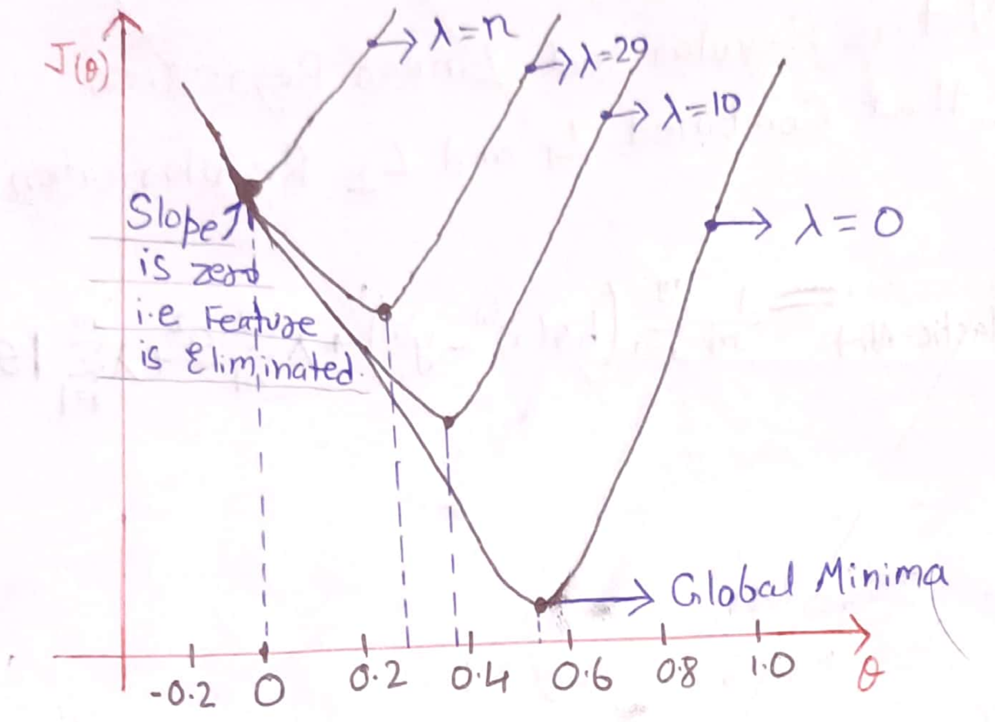
It is used when we have more features because it automatically performs Feature Selection.

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

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

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

*) Relationship Between λ 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.

∴ $\lambda\uparrow\uparrow$ θ_3 tends towards Zero and Finally θ_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.

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

⇓

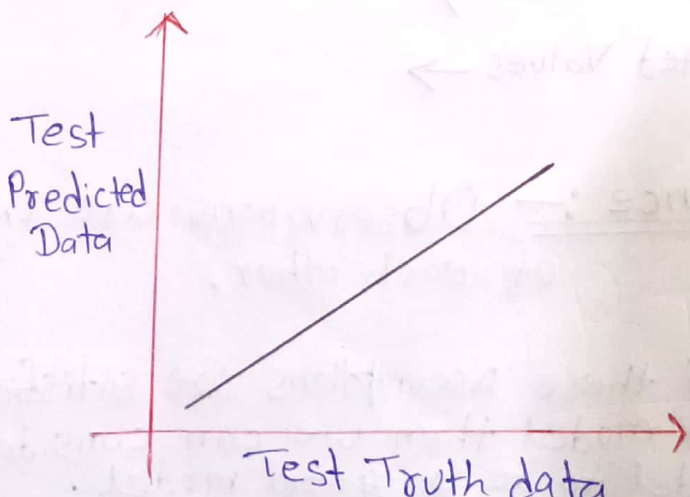
a) MSE (c) RMSE
b) MAE (d) Huber Loss.

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

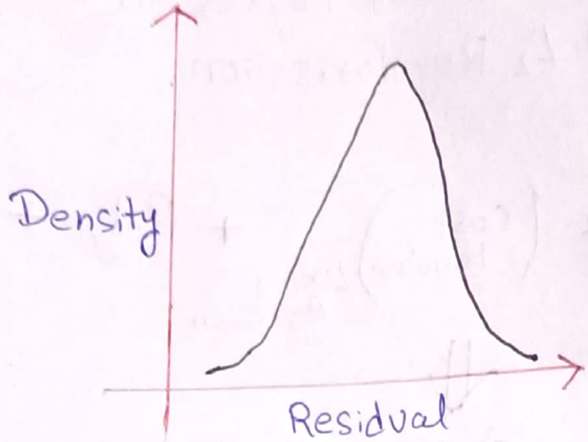
∴ In terms of handling bias, Elastic-Net is Considered better than Ridge and Lasso.

☆ Assumptions in Linear Regression :-

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

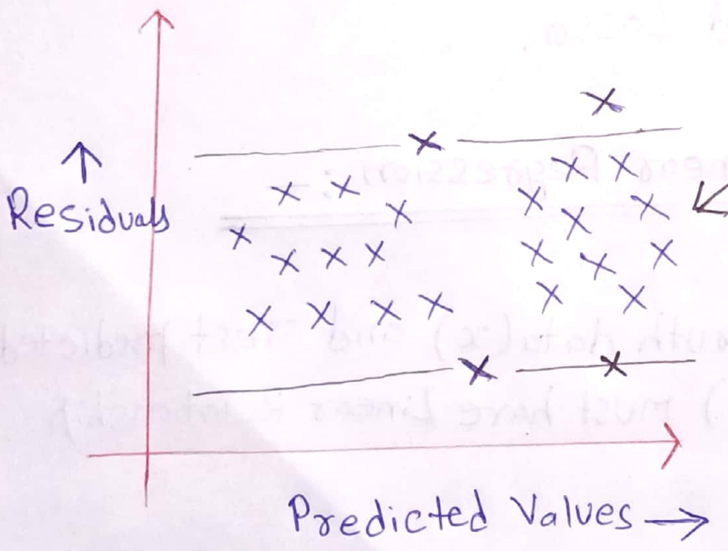


b) Normality :- Residual Plot Should be Normally Distribute



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.