Machine Learning
DAY-3 DATA SCIENCE

Regression

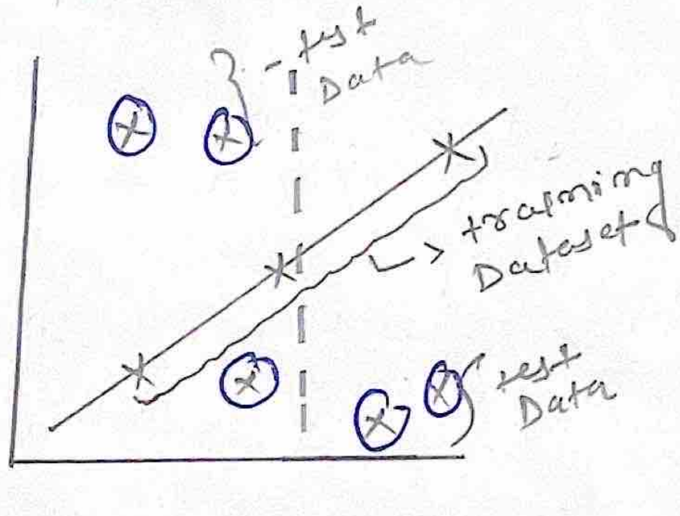
1. Ridge Regression (L_2 Norm or L_2 Regularization)

⇒ It is Model tuning method. that is used to analyse data that suffers from multicollinearity This Method Performs L2 - Regularization

AIM = To Reduce over fitting

Example
Lets Consider a scenerio where we have our train Data that has overfitted best fit Line

x = Training Data (Low Bias)
⊗ = Test Data (Low/High variance)



when
cost fn = 0
* overfitted Model

→ Now to overcome this overfitting situation we create another line with some error. for this we use Ridge Regression

*** Cost Function in Ridge Regression**

Cost fn :- $(\text{cost fn})_{LR} + \lambda \sum_{i=1}^{m} (\text{slope})^2$

$h_\theta(x)$
⇓
$\theta_0 + \theta_1 + \theta_2 + \theta_3$
↳ slope

λ = hyperparameter tuning

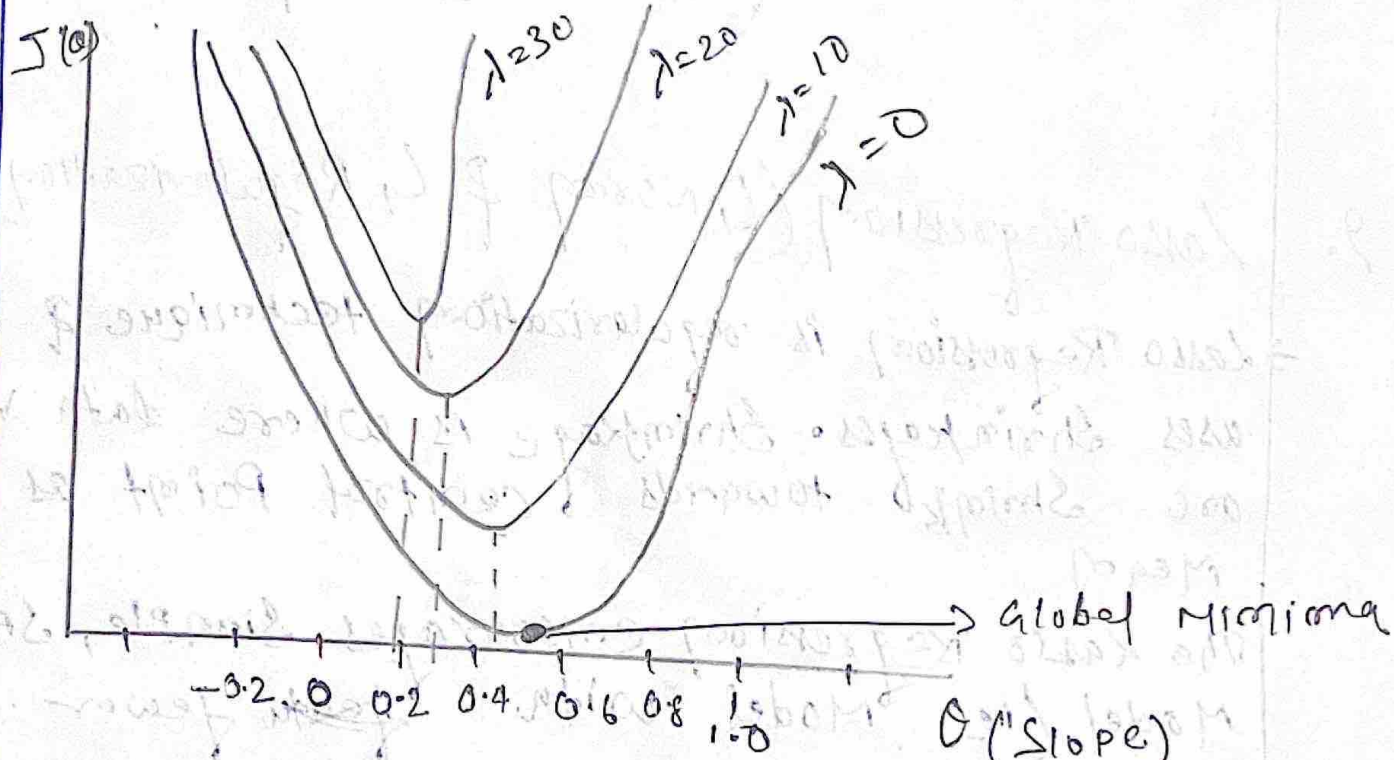
slope = slope of individual feature indipendent

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

$$\sum_{i=1}^{m}(\text{Slope})^2 = \theta_1^2 + \theta_2^2 + \theta_3^2 + \cdots + \theta_m^2$$

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

* Relationship B/w λ & Slope



→ From above Graph it is evident that as λ increase Slope decrease

→ λ is inversely Proportional to the Slope

ie : $\lambda \propto \frac{1}{\text{Slope}}$

At Worst Case (Cost Function) = 0 (Linear Regression)

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

$= \lambda(\text{Slope})^2$ — +ve, +ve

Cost fn (ridge) = +ve value

* There will be never be overfitting.

Note

In Ridge Regression slope (θ) value will reduce but it will never be reach zero

Since if θ reach zero $\Rightarrow$ feature will be deleted.

2. Lasso Regression (L_1 Norm & L_1 Regularization)

→ Lasso Regression is regularization technique & it uses shrinkages. Shrinkage is where data values are shrunk towards central point as mean

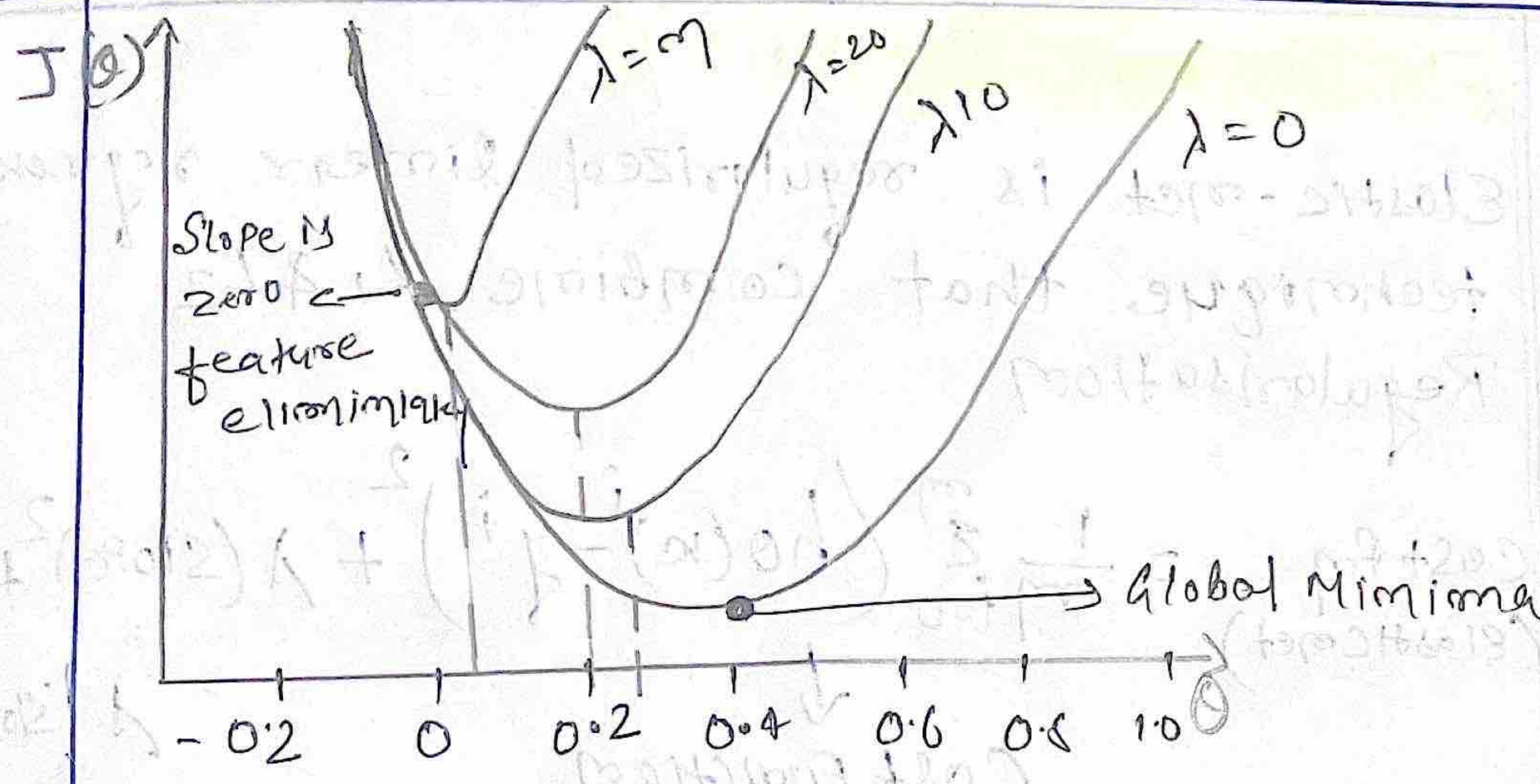
The Lasso Regression encourages simple, sparse model (i.e model with ~~featu~~ fewer feature

This model is used when there is high multicollinearity or when we want to automate certain part of model selection like variable selection/parameter elimination

It is used when we have more features because it automatically perform feature selection.

AIM → To reduce feature i.e feature selection

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



Example :- least correlated feature get eliminated

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

here x_3 is least correlated feature

as λ increase least θ_3 decrease to zero & finally it will become zero & the feature will get deleted

→ For outliers we must use lasso Regression.

3. Elastic Net Regression :-

Elastic-net is regularized linear regression technique that combine L_1 & L_2 Regularisation

$$\text{Cost fn (elasticnet)} = \frac{1}{m}\sum_{i=0}^{m}\left(h\theta(x)^i - y^i\right)^2 + \lambda(\text{slope})^2 + \lambda|\text{slope}|$$

↓

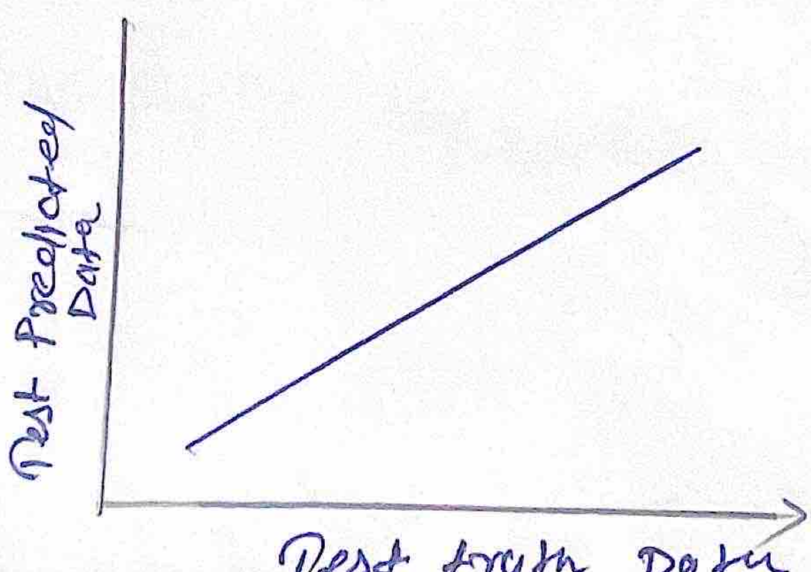
Cost Function May be

1. MSE
2. MAE
3. RMSE
4. Huberloss

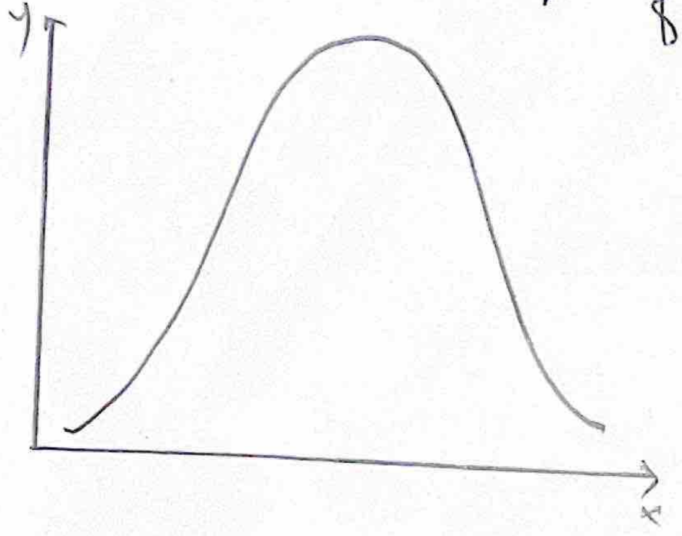
In the terms of handling bias, Elastic-net is considered better than Ridge & Lasso.

Assumptions in Linear Regression

Linearity :- The test truth Data(x) & Test Predicted data y Must have Linear Relationship



Normality :- Residual should be normal Distributed if Ploted

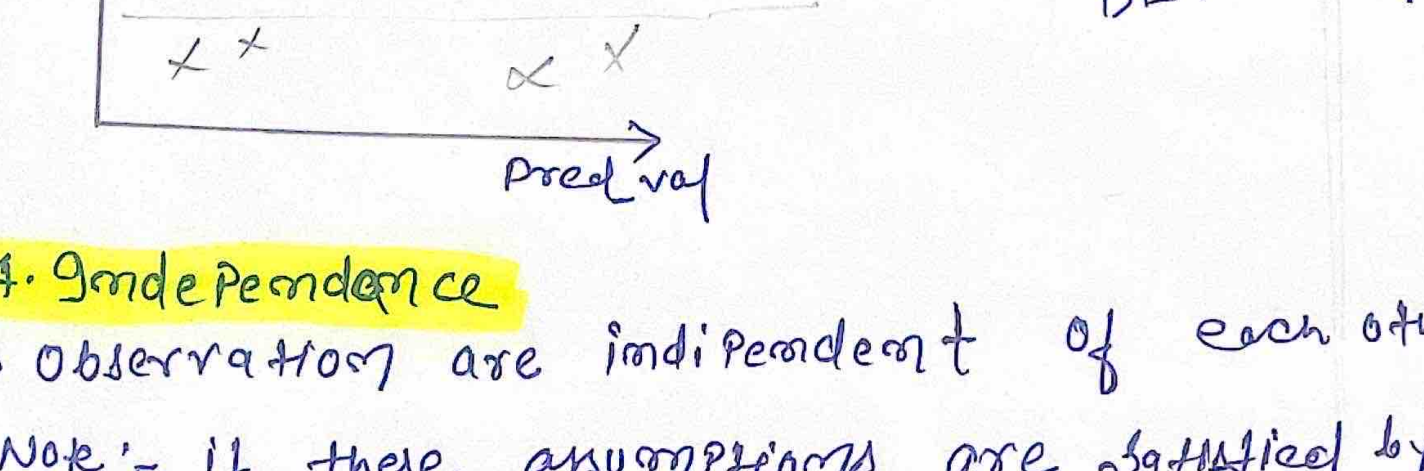


3. Homoscedasticity

~~The value of any~~

The variance of any value of Residuals is same for any value of X.

Scatter Plot of Prediction Must have uniform Distribution ~~of~~ with Residuals



4. Independence

→ observation are indipendent of each other

Note :- if these assumptions are satisfied by own Model then we can consider our Model to be Good Model.