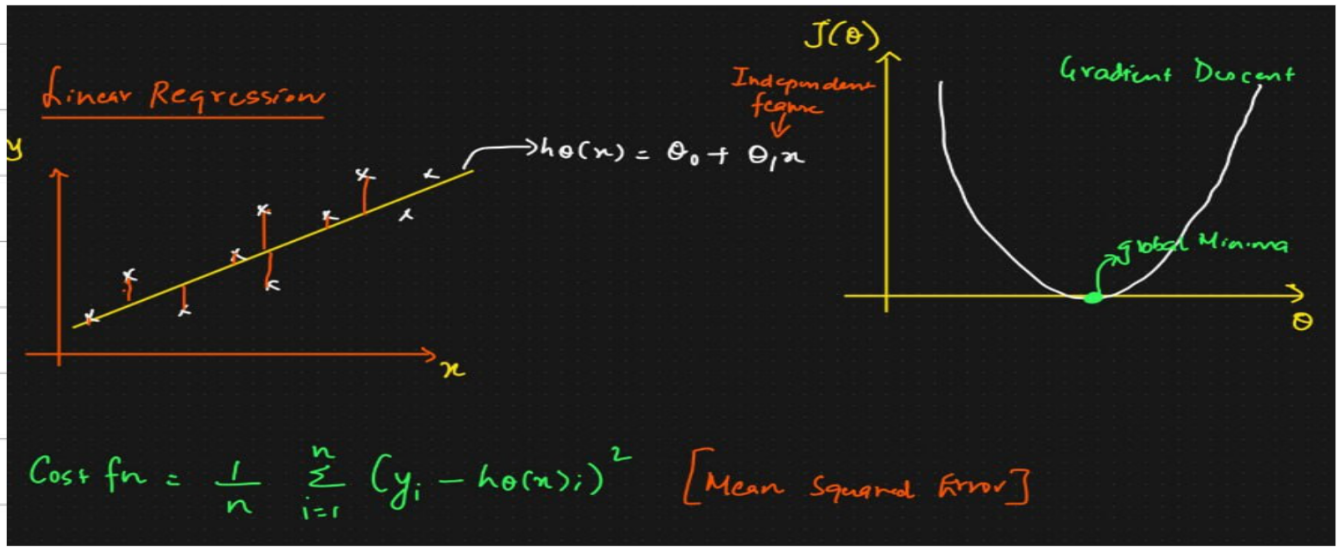
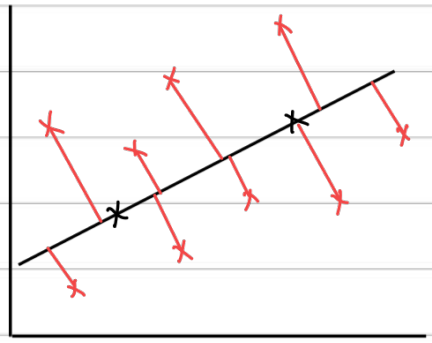
Ridge, Lasso, Elasticnet Regression

We know,



1)- Ridge Regression (L2 Regularization)

Reducing Overfitting.



x → test data points

x → train datapoints

$$\text{Cost fun}^n = \frac{1}{n} \sum_{i}^{n} (y_i - h_\theta(x))^2 = 0$$ for training data points

&

$$\text{Cost fun}^n = \frac{1}{n} \sum_{i}^{n} (y_i - h_\theta(x))^2 \neq 0$$ for test data points

∴,

there for this situation is called "Overfitting", where

Train dataset → Accuracy ↑↑(high) (low bias)

Test dataset → Accuracy ↓↓(low) (high varience)

So, to overcome this overfitting we use ridge regression.

∴,

in Ridge Regression:

Our aim is to make the best-fit line to not fit the training line perfectly.

thus, we add 2 parameters to the cost function such that:-

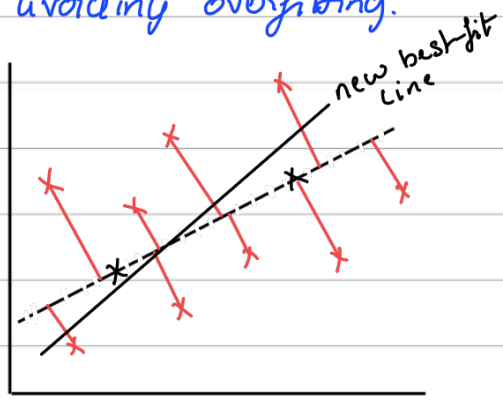
$$\boxed{\text{cost function} : \frac{1}{n} \sum_{i}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{n} (\text{Slope})^2}$$, λ: hyperparameter

using above eq suppose, $\lambda = 1$

$\Rightarrow J(\theta) \cdot$ cost funⁿ $= 0 + (\lambda = 1)(\theta_1)^2 \Rightarrow$ penalizing cost function

So;

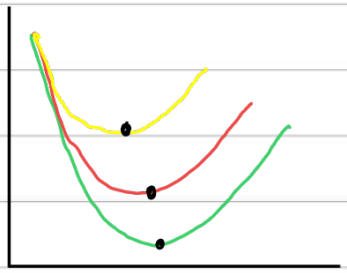
in the above case the best fit line will not fit the training set thus avoiding overfitting.



x → test data points

x → train datapoints

Hyperparameter (λ) and Slope (θ_1) relationship :-



Observations :-

① as λ ↑es Slope (θ_1) ↓es. ∴ ✱ $\boxed{\lambda \propto \frac{1}{\text{Slope}}}$

② as λ ↑es, global min shifts up

i.e λ↑es, Global min ↑es

Q) How the eqⁿs will chage for Simple and Multiple Linear Regression as Hyperparameter is introduced.

① Simple Linear Regression

- $h_\theta(x) = \theta_0 + \theta_1 x$
- Cost funⁿ $= \frac{1}{n} \sum_{i}^{n} (y_i - h_\theta(x))^2 + \lambda[(\theta_1)^2]$, ($\because \theta_1 =$ Slope)

② Multiple Linear Regression :-

- $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_3 x_3$
- Cost funⁿ $= \frac{1}{n} \sum_{i}^{n} (y_i - h_\theta(x))^2 + \lambda[(\theta_1)^2 + (\theta_2)^2 + (\theta_3)^2]$

2)- Lasso Regression :-

also c/a : "<u>L1 Regression</u>" → (feature selection)

$$\boxed{\text{Cost fun}^n = \frac{1}{n}\sum_{i}^{n}(y_i - h_\theta(x))^2 + \lambda \sum_{i}^{n} |\text{Slope}|}$$

Let's.

consider Multiple linear regression for this case, such that

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

and

Suppose after training the model we found that

$$\theta_0 = 0.52, \quad \theta_1 = 0.65, \quad \theta_2 = 0.72, \quad \theta_3 = 0.12$$

$$\Rightarrow h_\theta(x) = \hat{y}_i = 0.52 + \underbrace{0.65\, x_1}_{} + \underbrace{0.72\, x_2}_{} + \underbrace{0.12\, x_3}_{}$$

means that for a unit change in x_1 0.65 change will happen in y

means that for a unit change in x_2 0.72 change will happen in y

means that for a unit change in x_3 0.12 change will happen in y — ref 1

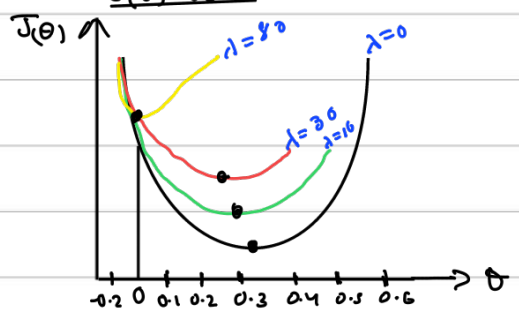
ref 1 :- in ref-1 as there is only 0.12 change in y ∴ it is considered to be very less correlated with y. So this feature can be removed. and for this process we us "<u>Lasso Regression</u>"

Now,

Suppose $\lambda = 1$

$$\Rightarrow J(\theta) = \text{Cost fun}^n = \underbrace{\frac{1}{n}\sum_{i}^{n}(y_i - h_\theta(x))^2}_{\text{Error}} + \underbrace{\lambda\left[\sum_{i}^{n} |\text{Slope}|\right]}_{1\,[|\theta_1| + |\theta_2| + |\theta_3|]}$$

$J(\theta)$ Vs θ :-



∴ as λ ↑es the slope/global minima starts to move and until is become zero such the θ_3 term is removed

i.e $\lambda \uparrow\uparrow$ Slope ↓

but at one point $\boxed{\theta \approx 0}$

$$h_\theta(x) = \hat{y}_i = 0.52 + \underbrace{0.65\, x_1}_{} + \underbrace{0.72\, x_2}_{} + \boxed{0.12\, x_3}$$

→ as soon as Lasso Regression is added this feature becomes zero.

3)- Elastic Net Regression:-

Elastic Net Regression is a combination of both Ridge Regression and Lasso Regression i.e

$$\boxed{\text{Elastic Net Regression} = \begin{cases} \text{① Ridge: Reduces Overfitting} \\ \quad + \\ \text{② Lasso: Feature Selection} \end{cases}}$$

$\therefore$

$$\boxed{\text{Cost fun}^n = \underbrace{\frac{1}{n}\sum_{i}^{n}(y_i - h_\theta(x)_i)^2}_{\text{MSE}} + \underbrace{\lambda_1 \sum_{i}^{n}(\text{Slope})^2}_{\text{Reduce Overfitting}} + \underbrace{\lambda_2 \sum_{i}^{n}|\text{Slope}|}_{\text{Feature Selection}}}$$

, λ_1, λ_2 : [Hyper parameter tunning]