RIDGE REGRESSION

- Also known as L2 Regularization. / L2 norm.

★ Aim of R.Reg. → Reduce Overfitting

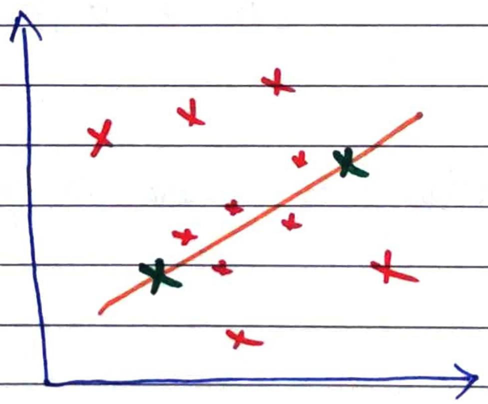
- How it Reduce Overfitting?

Let say in our **Training Data** we have 2 datapoints

In 2 datapoints, our best fit line will pass perfectly through them. (Perform exceptionally well)

↳ That means data is OVERFITTED.

Our Cost function will be exactly 0.



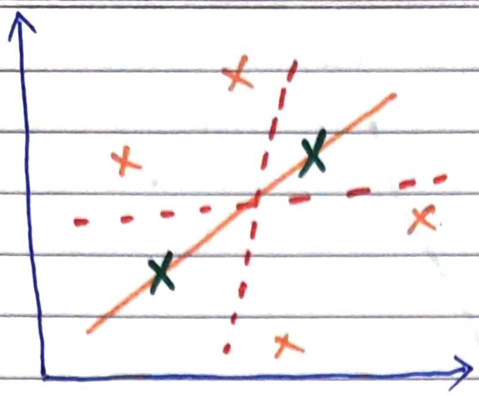
Overfitting

x Training Data : Low Bias

x Test Data : Low or High Variance

It is always a good thing that you should not try to overfit the data.

↳ In order to prevent this we will create a line....



We create a best fit line in different way (- - -)

This will reduce the error with test data.

In Linear Regression we are always going to get the BEST FIT line

But to create - - - line we use Ridge Regression which will reduce the error in our Test Data.

In order to get this line, we will use the ....

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

$\because \lambda$ = hyperparameter

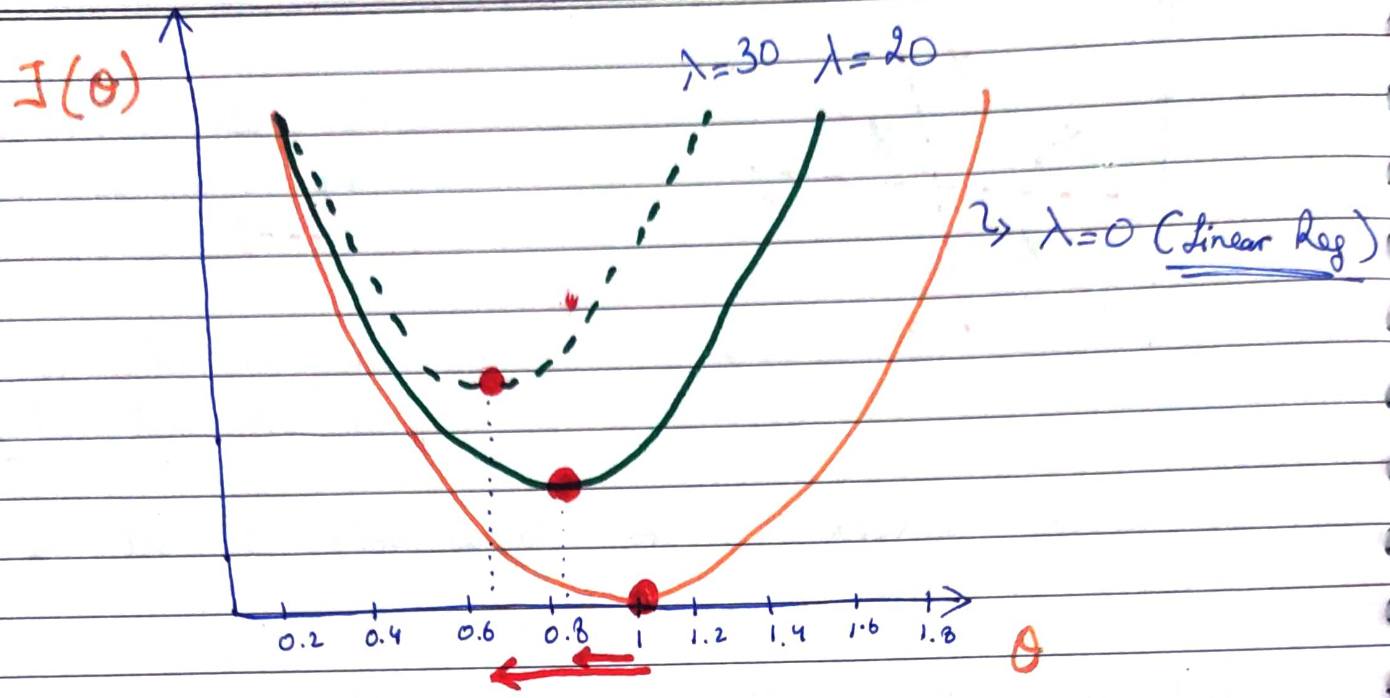
slope = θ_1

If ~~at~~ we have multiple slopes we square their summation
$\lambda(\theta_1 + \theta_2 + \dots \theta_n)^2$

whenever $\lambda = 0$, we will get Linear Regression

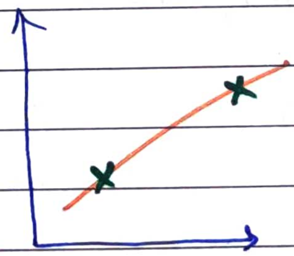
** *** Q Relation b/w θ (slope) and λ?

A: As we saw, when our $\lambda = 0$, we get Linear Regression. That means we will get GRADIENT DESCENT curve.



Observations :-

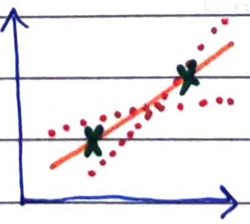
- Slope value decreased when Lamba (λ) increased. Because we got new Global Minima on left side.
- Considering overfitting in mind, when we decreasing slope value, its minimizing cost function.
-



Initially, when $\lambda = 0$
Cost fn $= 0$

Adding λ & slope,
↳ Cost fn $= 0 + [+ve]$

Since my cost function get weighted, the line will change.



It may go up or down depending where Best fit will fall it.

- λ is inversely proportional to slope.

- When θ value changes with new weight with λ, a new Best fit gets created.

- In this way it **NEVER** get **OVERFITTED**

★ So with Ridge Regression, you know it will never Overfit!

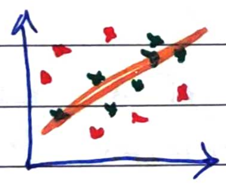
Recap:-

- Let say we have a problem statement which is Overfit.

  Overfit Condition →

  Training Data → High Accuracy = 95%

  Test Data → Low Accuracy = 60%



- In order to solve this problem, we use Ridge Regression. Because in Ridge Regression we don't only just focus on Cost function, also we add something to it.

$$\text{Cost fn} + \lambda(\theta)^2$$

- Adding λ value reduces θ value since our Global Minima gets shifted on left side.

- We keep on changing λ value until we get result.

- θ will never be Zero!

Lasso Regression

- Also known as L_1 Regularization / L_1 Norm.
- Aim → Reduce the features.
  ↳ feature selection.

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

- How its done...
  - let say I have an initial Linear Regression where $\lambda = 0$
  - When I increase my $\lambda = 10$, we get a new curve which shifts global Minima & decreases θ (slope).
  - As you keep increasing λ value, we see that a point comes where our ~~the loss~~ global Minima will stuck at 0.
  - Some of coefficients become zero.

g. $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$ {3 Independent feature}

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

we can see if there is 1 unit movement in y axis we have 0.54 unit movement in x_1 axis.

That means x_3 is least correlated feature.

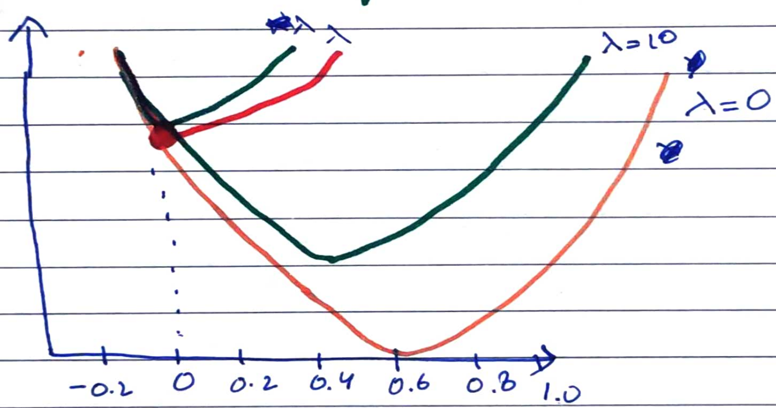
With hyperparameter tuning, increasing λ, this weak feature becomes zero.

↳ feature deleted!

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

$$= \underbrace{\boxed{\theta_0 + 0.54x_1 + 0.25x_2}}_{\text{Important Features}} + \underbrace{\boxed{0.10x_3}}_{0}$$

Removed



- Maximum point θ goes till zero!
- Whichever features are NOT HIGLY CORRELATED, they become zero and they get deleted.

★ When your dataset has outlier, use ~~Ridge~~ LASSO

ELASTIC NET

- Combination of Ridge & Lasso [L_1 & L_2 Norm]

- Cost fn $= \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x)^i + y^i)^2 + \lambda_1 (slope)^2 +$

$\lambda_2 |slope|$