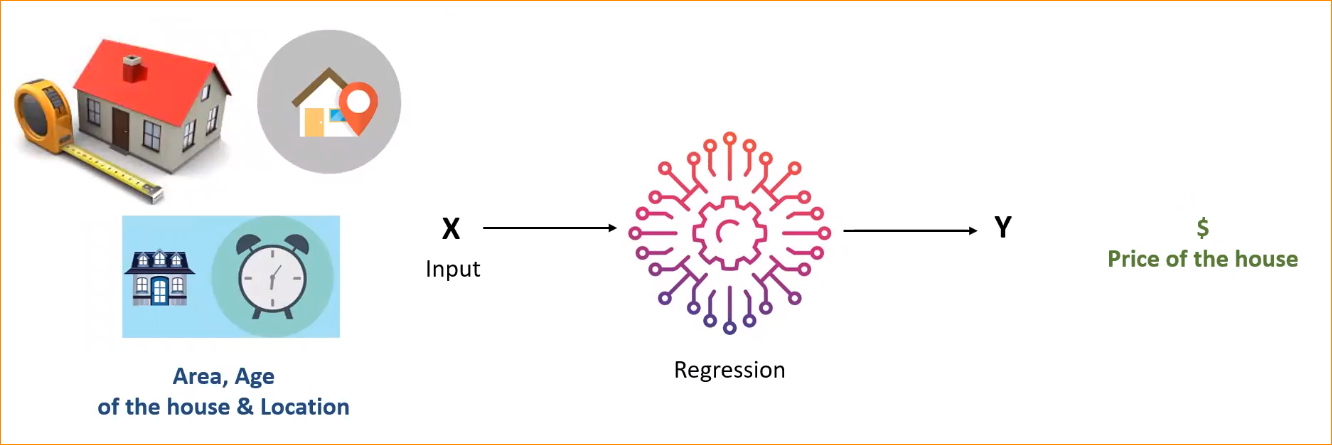
Multiple Linear Regression

That mean much more independent variable and rest will remain the same.

We still assume the relationship is linear.

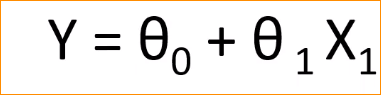
Ex –



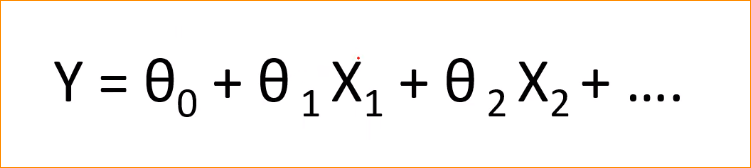
Depend on the given features – predict the price (Outcome, Dependent variable) of the house.

* Multiple independent variable – X1, X2, X3 …
* Relationship between the independent variables and Y is assumed to be linear.

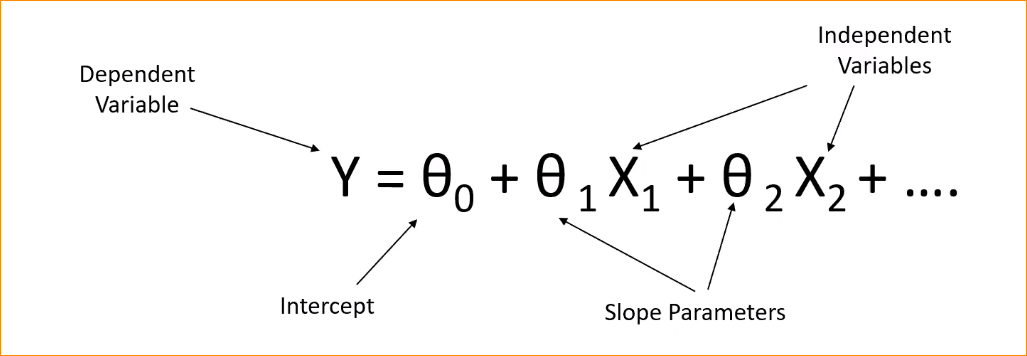
Simple Linear Regression equation –



Multiple Linear Regression equation –



When you add new variable – X2, X3 …

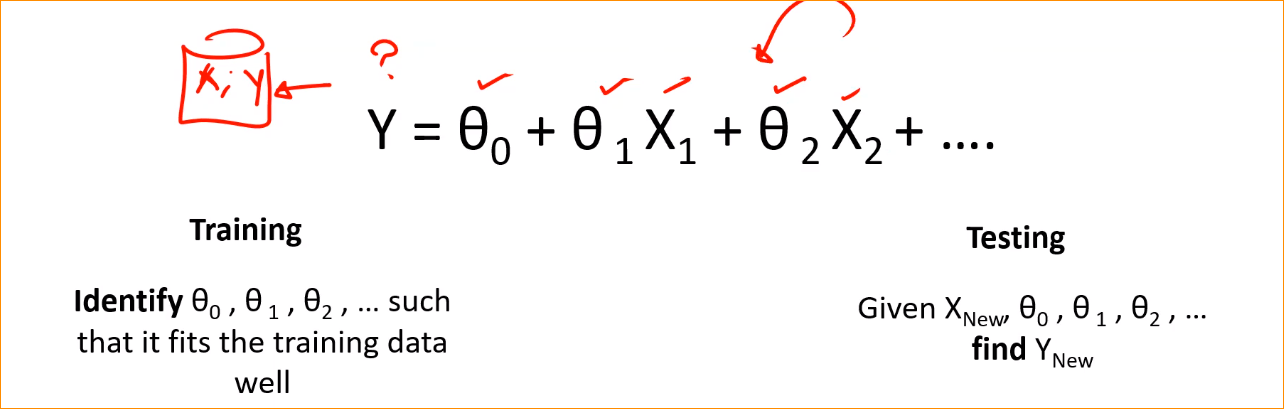


Linear Regression with Two Variables-

Plot – It will be a plan

So if you have **k independent** variables – plot will be k+1 dimensional space (Can not be visualize)

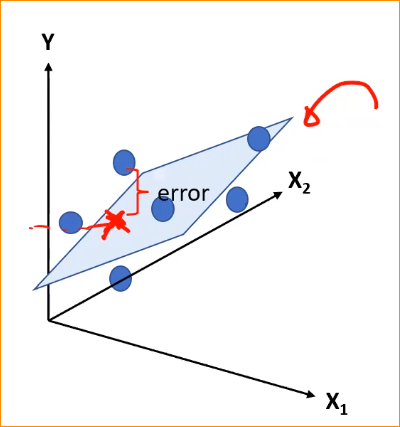
How to solve this problem?



* Initially We have Training dat (Collection of X..)
* Corresponding Y (It will be a single value)
* In the training we will try to recognize the parameter. (theta 1 , theta 2, theta 3 … ) those are should fit our training data.
* When the training is done, in the real world when you go and apply your model, when a new bunch of X come

**Error in prediction –**

When, theta 1, theta 2, theta 3 … ready when you apply to the equation and get Y that will be somewhere on the plan. When we project X1 and X2, can get Y. But Actual is a different one.

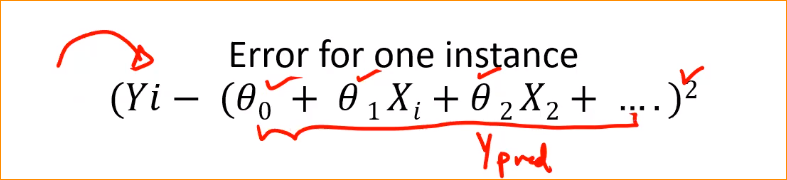


What we will do during the training process, We will some all the error and try to minimize it. So we will fine the minimal values of thetas to minimize the error.

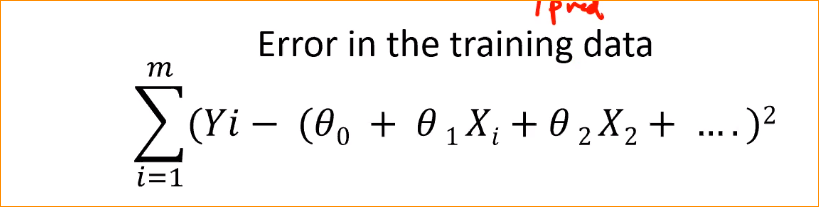
Objective – Minimize the error for all the m points in the training data.

Let take single instance-

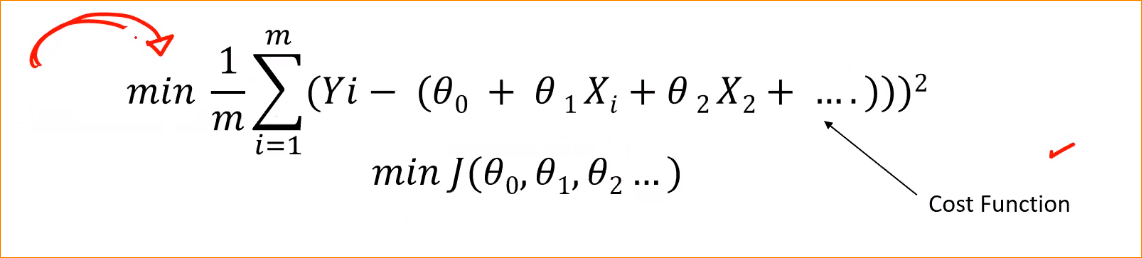
We have Yi , So according to X can get Ypredicted – So can measure error, here get Square of it.(all plues)



So We will get sum across around data all data. Each training data we made a square error and we summit it.



And for the normalization, we will divide it in to m. that is our cost function.



So objective is figure out the thetas.

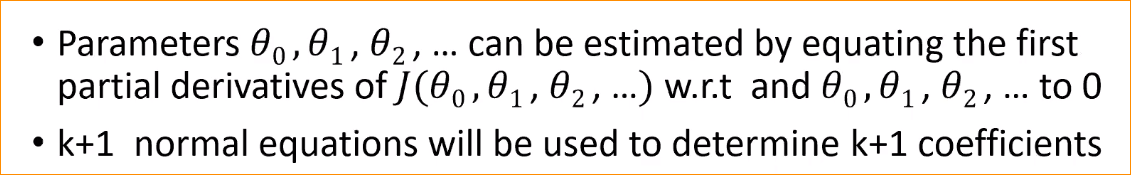
Such that the equation is minimize.

That mean the total training error will be minimize.

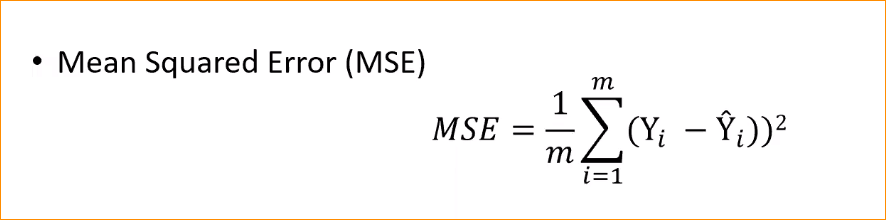
Can use the same technique to estimate to parameters.

We can get partially derivative respect to the each parameter.

Equals to 0.



Evaluation Metrics (Check the model performance)



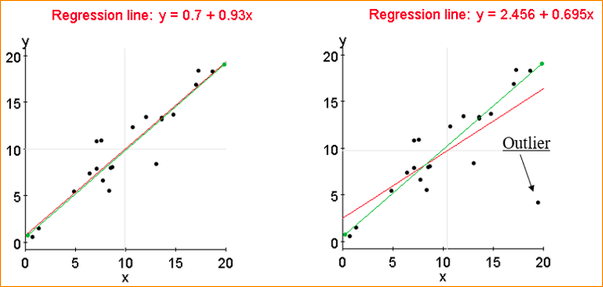
In data we have Xi and Yi

So use that and found MSE

MSE use for the evaluation Metrix

Why MSE is bad ?

It is suspectable for outliers –

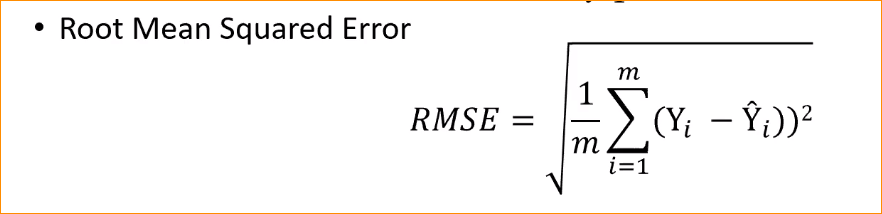


Lets get only one variable- We have this pattern , Can draw a nize line, but there is some data those not match with that pattern (Far away in the graph) So we call outliers.

There have several techniques to remove the outliers

If have outliers , Can see the error will be really large. The when we take a Square of it. Its going to be too large. So we have to adjust the theta a lot. Because the total error should be minimized.

For that there is an alternative.



Normally the MSE could large value, good to get RMSE – Roor mean squared error

So here this is effective for the outliers.



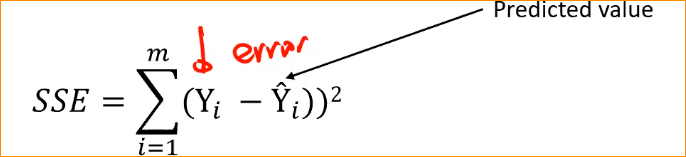
Just get error and normalize it

This is very stronger with outliers.

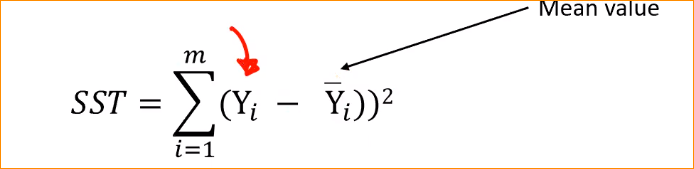
**R Square error =**



Sum of Error



Variation of mean –



If you have the best model in the world-

Yi – Actual one

Yi^ - Prdicted one

Those two should equal 0 that nean 1 -0

R squre value = 1

One SSE and SST equal each other (two value – predict and actual value close each other)

That mean

R squre value = 0 🡪 this will be a use less model

When R^2 close to 1 🡪 that mean good value

If SSE > SST what happen – R^2 “-“ value that mean you make larger error. This mean you are making very Wrong Prediction.

So the Max R^2 value is 01

