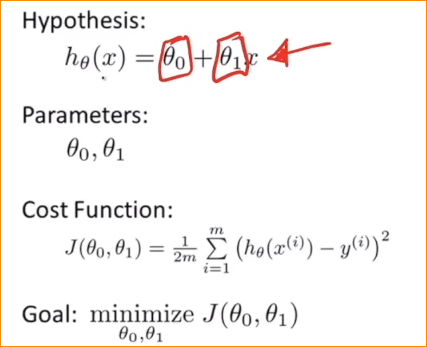
**Logistic Regression 1**

**gradient descent**

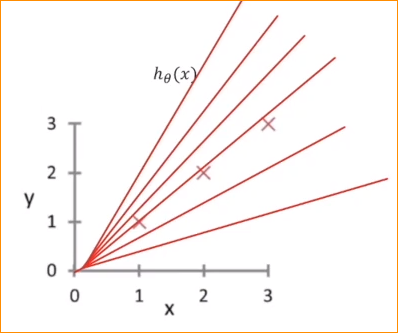
If we consider linear regression –



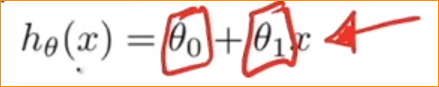
We have to do find best theta 0 and theta 1 to find the best fit model.

Example – in here we have 3 data points-

We can draw millions of regression lines among them what we have to find is what is most suitable regression line.

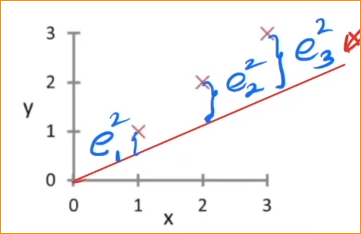


To adjust we have intercept and slope. (theta 0 and theta 1)



To choose best regression line. We use cost function. That mean minimizing the error.

What is error mean?

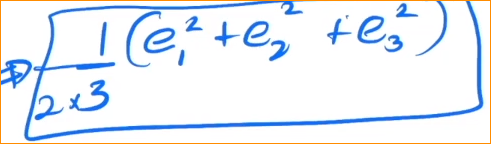


In here if we consider h(x) = 0.7X

Now I draw the regression line and get the error squre of each and every data point.

Get total all errors and devide number of points. Devide 2 (only use deferantiation)

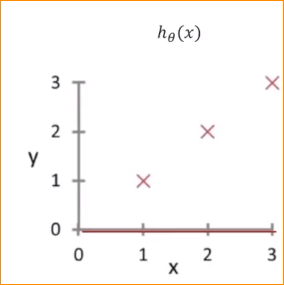
So our object is to fined theta 0 and theta 1 to minimize this error.



To fined the most sutable theta 0 and theta 1 , while minimizing the cost function – that is we are using **gradient descent** method.

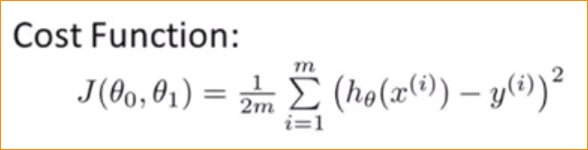
Example 02 –

In here there is no error (because line going each and every data)- So our squre error cost function (J theta) is going to 0.



So note – 99.99% models the cost function is not going 0.

Now

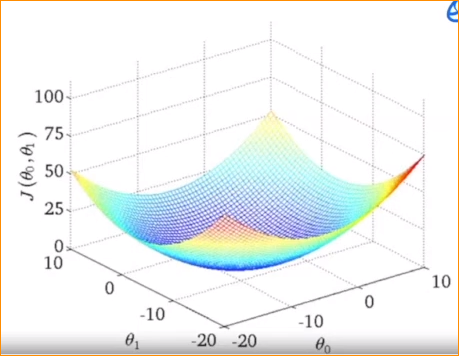


To minimize the cost funtion we use gredient decent algorithm.

* Let's see how to find the smallest colst values and coefficients

Just we are making the graphs that change the values-

|  |  |  |
| --- | --- | --- |
| **Theta 0** | **Theta 1** | **J(theta 0 ,theta 1)** |
| 0.1 | 0.1 | 1 |
| 0.1 | 0.2 | 1.5 |
| . | . | . |
| . | . | . |



Objective – what are the minimum cost value and to that what are the theta 0 and theta 1.

So in the gredient decent – just we want to select any point of this graph.

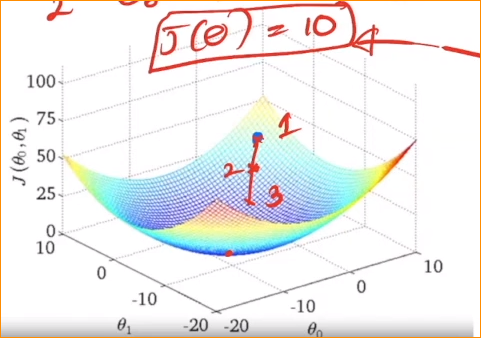
Lets say that values are –

Theta 0 = 5, theta = 8 and J(theta 0 , theta 1) = 29

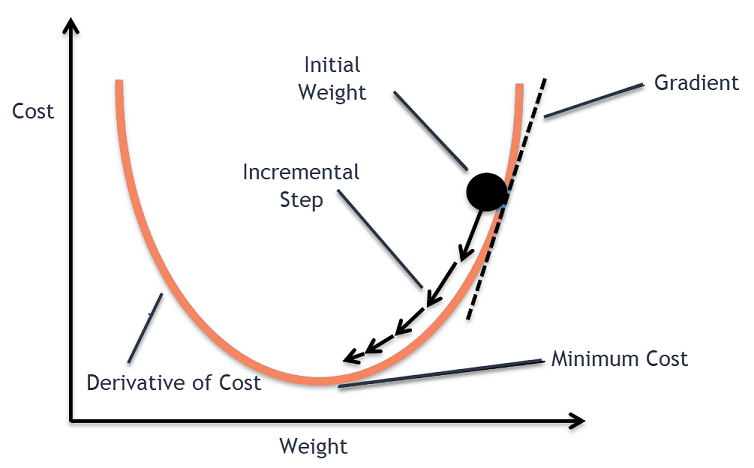
In graph we are always tring to go down.

If you come minimum point mean minimum value of cost function.

Just think at the moment cost function is 0.1



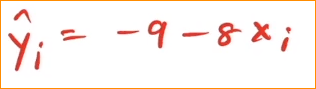
If you reach this point mean if you go around mean it increase the cost function



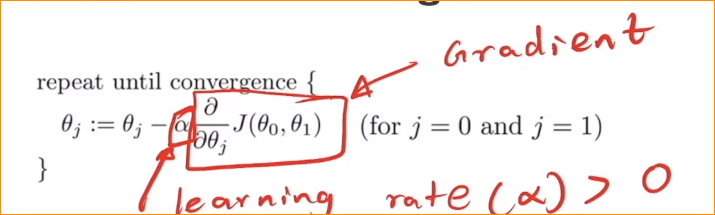
1. In this minimum point, need to calculate the cost function
2. In this point what is the theta 0 and theta 1

Lets assume theta 0 = -9 and theta 1 = -8.

Those are the most suitable value to represent our model.

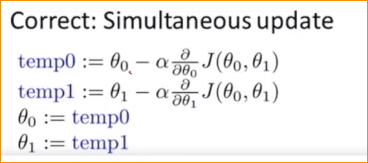


To do that we have to get partially derivative of cost function.

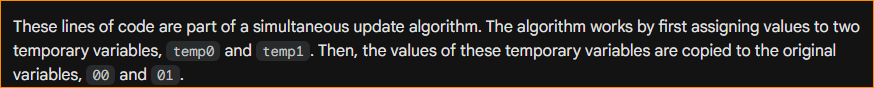


In here alfa is learning rate

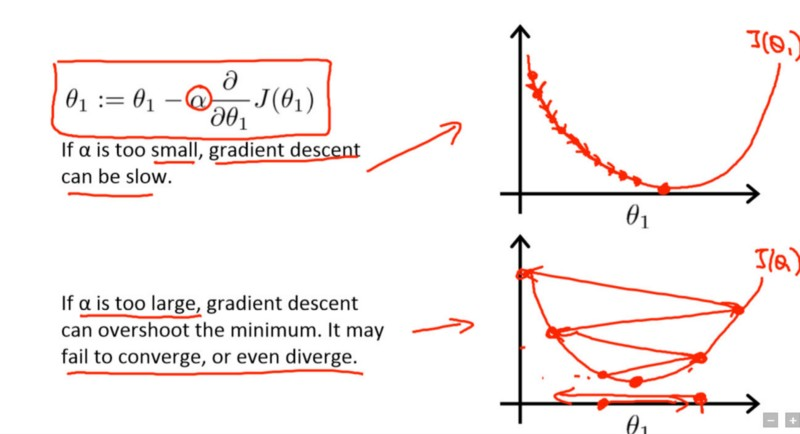
This is hyper parameter that you have to define. And this should be positive value.



Note –



In here this alfa mean how long you are expect to move down





* If we define small alfa mean it going very short steps. It will take long time to find the lowest point.
* If you take very large steps, it maybe over shoot the minimum point.

So we have to vary careful to define alfa.