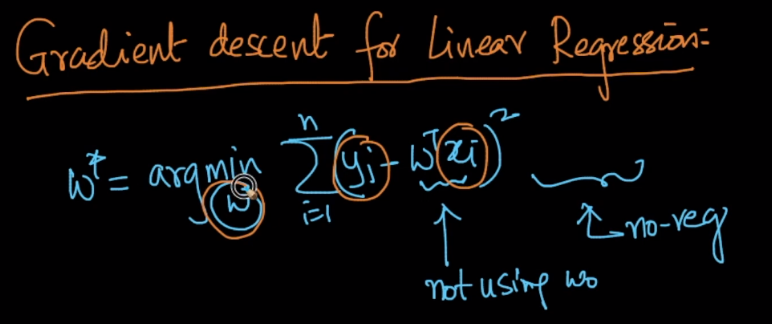
Let’s apply gradient descent for linear regression.

Objective function for linear regression was as given in below image.

Here note, that we are finding optimal w and not x.



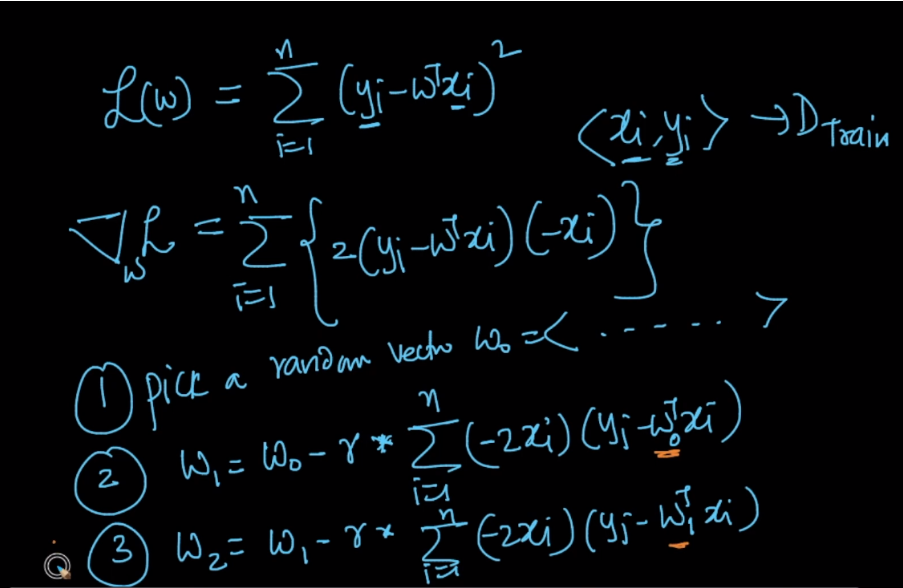
So for objective function, it’s derivative value is given below.

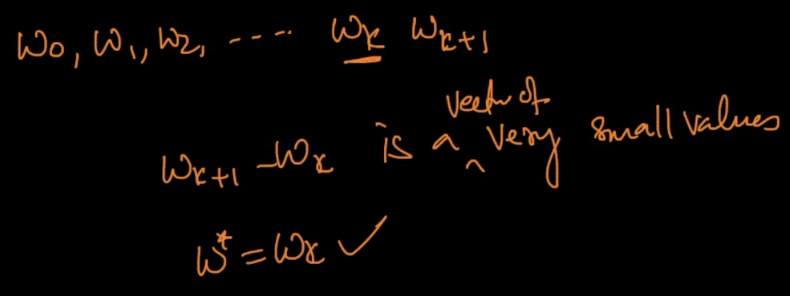
So we do following things:

1. Pick a random vector w0.
2. Calculate w1
3. Calculate w2

Repeat until difference between new w and old w is very small.

The w, where the difference is very small is said to be optimal w.





**Limitation of gradient descent:**

But as we can see, at each time we are calculating new w, we’ve to perform computation for ‘n’ datapoints,

If n is very large then computing this sum will be very expensive in gradient descent.

So for this reason we use **stochastic gradient descent.**

