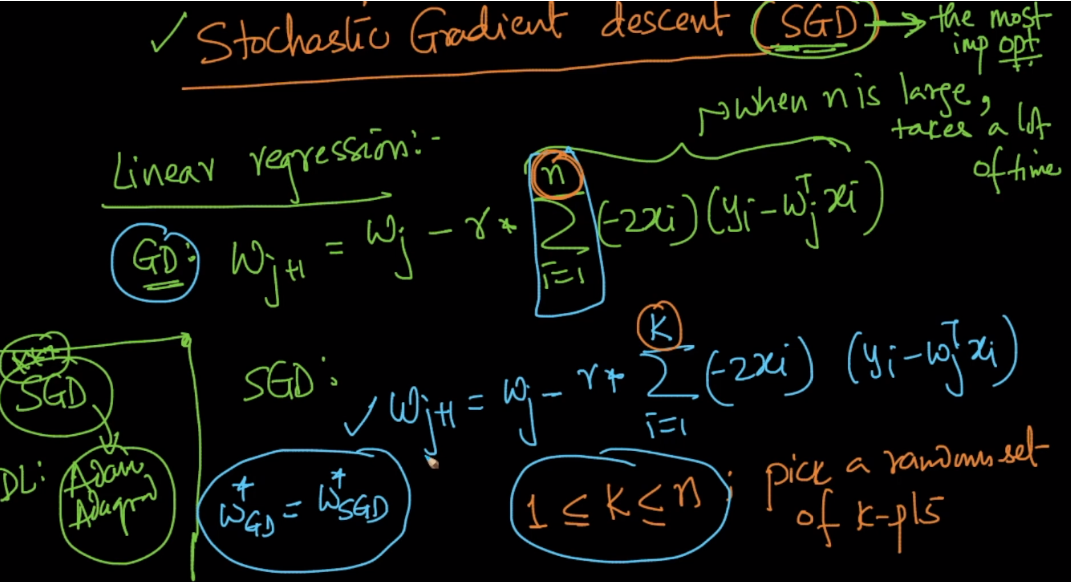
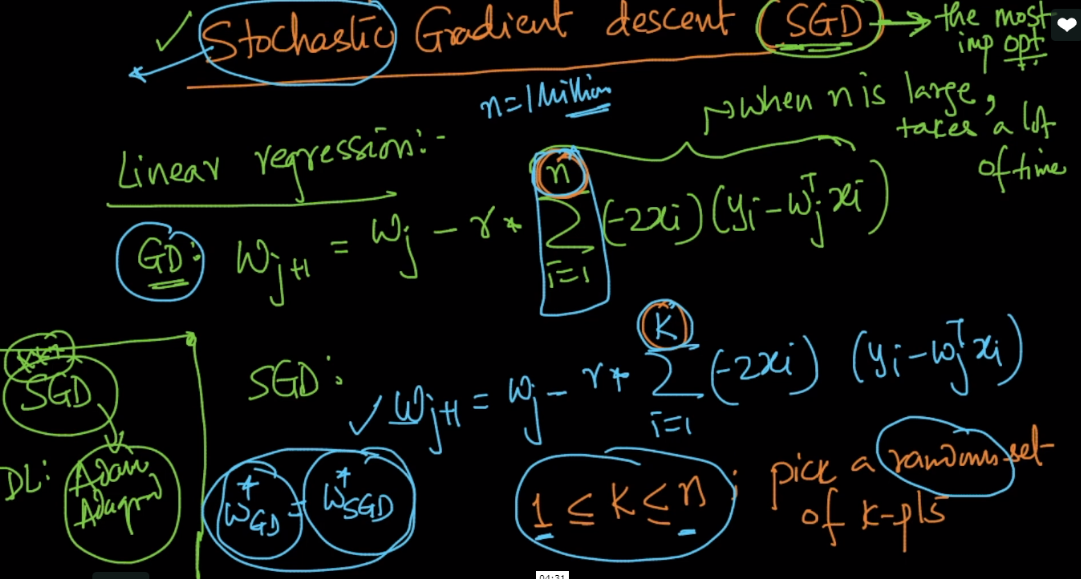
**Stochastic Gradient Descent:**

Let’s say we’ve ‘n’ data points where n is very large, Now in stochastic GD we don’t use all ‘n’ points, we use only **k** **random** points, where 1 <= k <= n.

**Note:** For each iteration of finding new **w** we always pick different K random points.

And by doing this we’ll get the optimal w equal to the optimal w we obtain in normal GD.





In SGD as no. of K decreases the more no. of iterations needed to optimal w.

But even more iterations are needed as compare to GD, it takes less time than GD.

