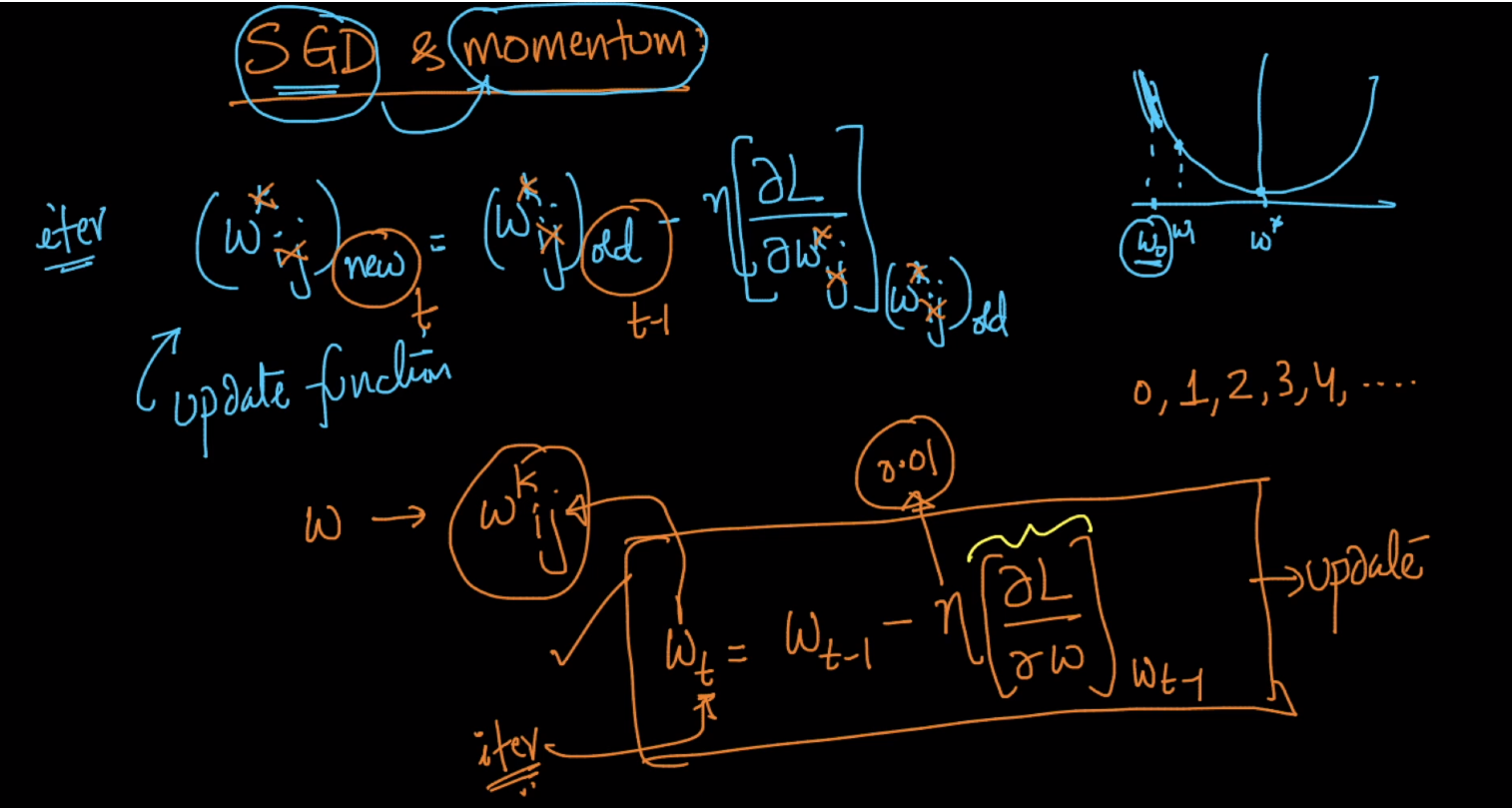
**SGD Recap**

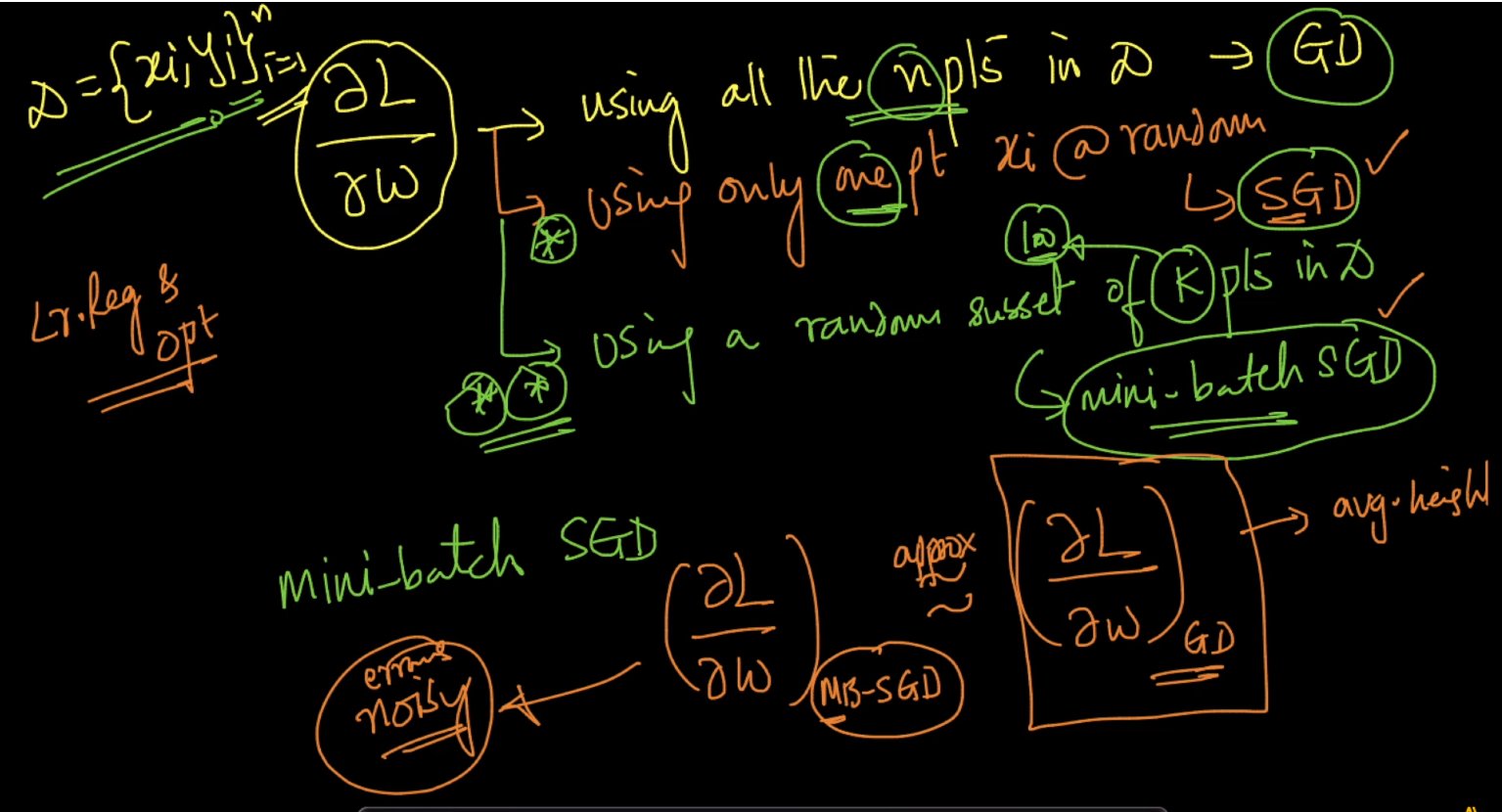
Below image shows about gd we already know about this.



Now we can find out gradient using gd, sgd, mini-batch sgd.

In dl we use sgd or mostly we use mini-batch sgd.

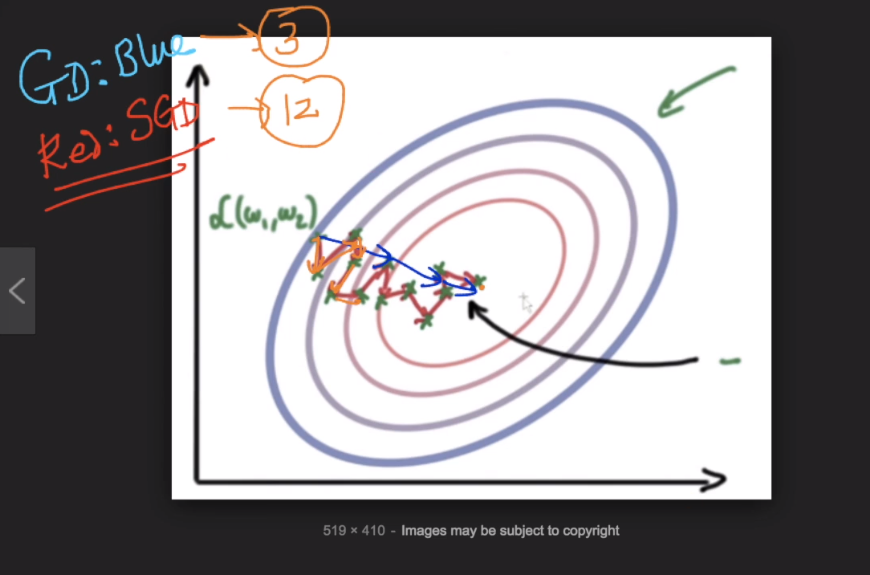
In mini-batch we got approximation of gd not actual gd therefore if we use mini-batch sgd we got noisy updates.

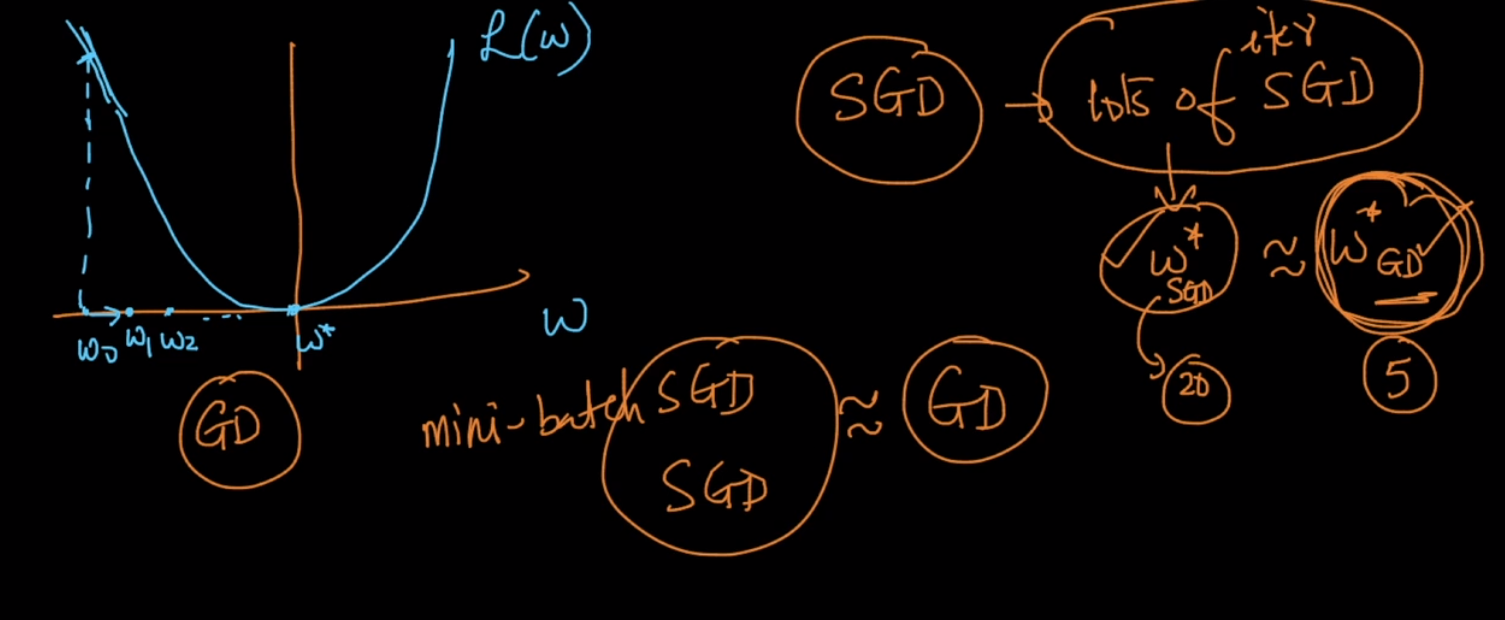


Below contour shows about gd

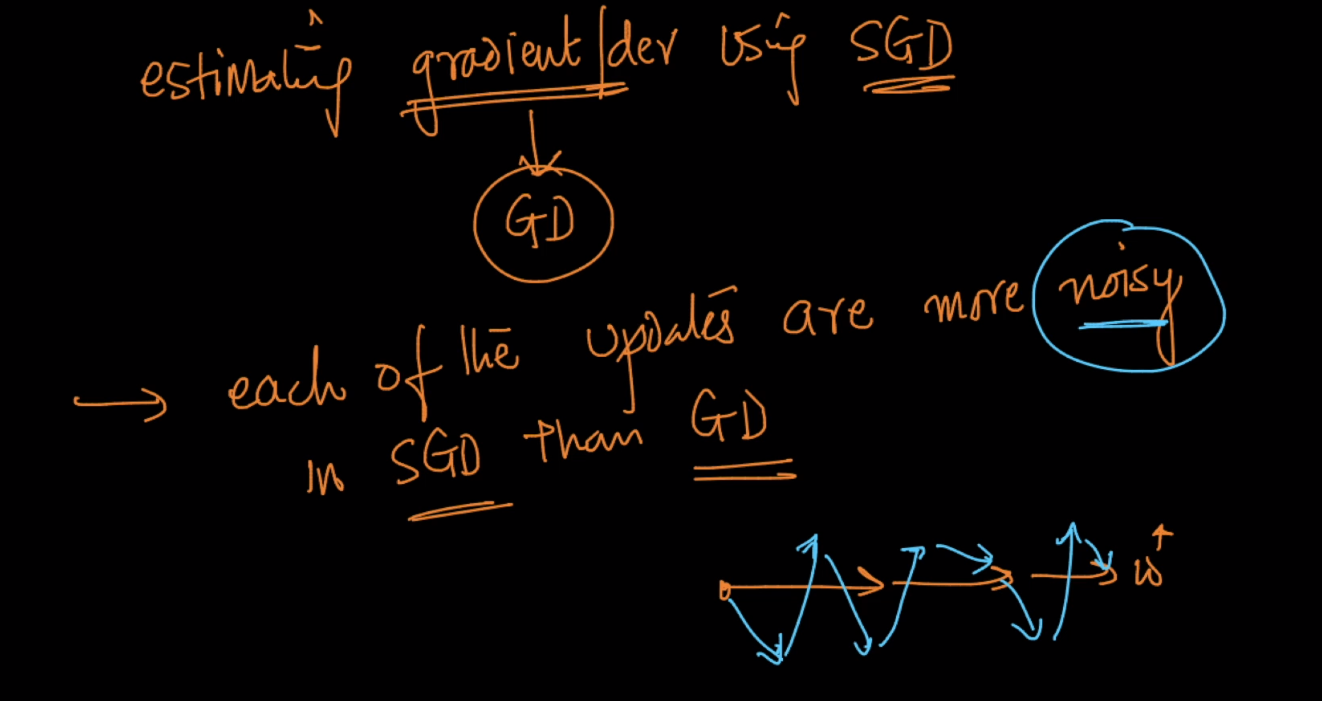


Below is the countour of sgd in which blue shows gd and by this we converge faster in 3 iteration and as mini batch sgd or sgd is approximation of gd and it have some noise therefore it converge slowly and take more iteration 16 to converge as shown in below image.



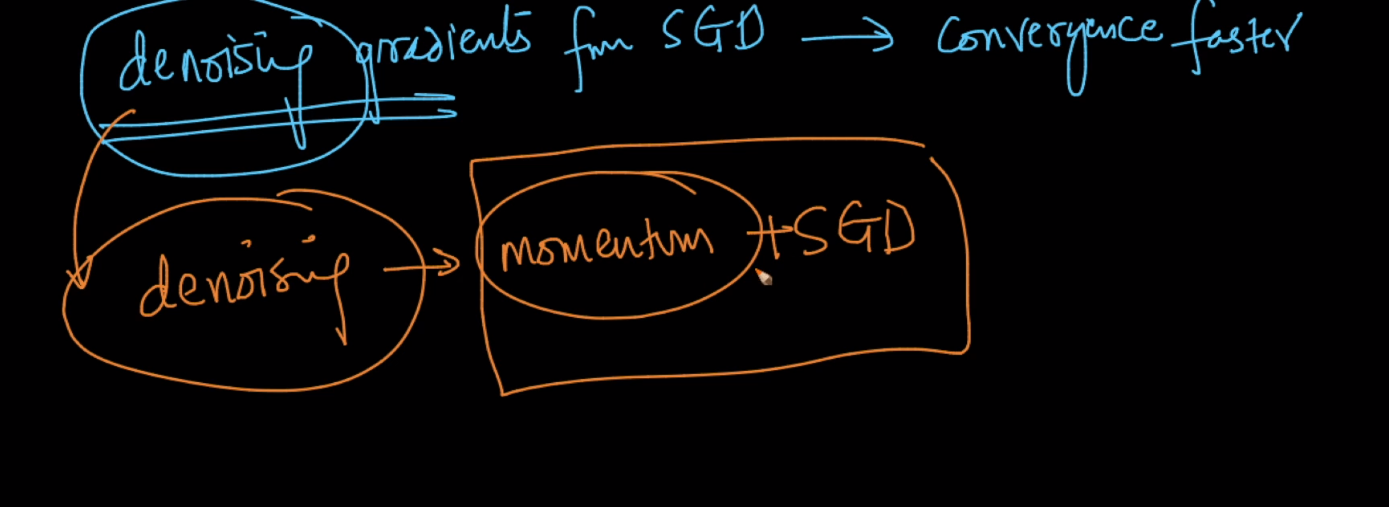


So if we are estimating gradient/derivative using sgd then each of the updates are more noisy as compared to gd



So to avoid this problem we use denoising gradients for sgd and it helps in convergence faster.

Denoising concept helps in momentum will discuss later.



Comments :

