

We read various optimizer so question is which optimizer to use when?

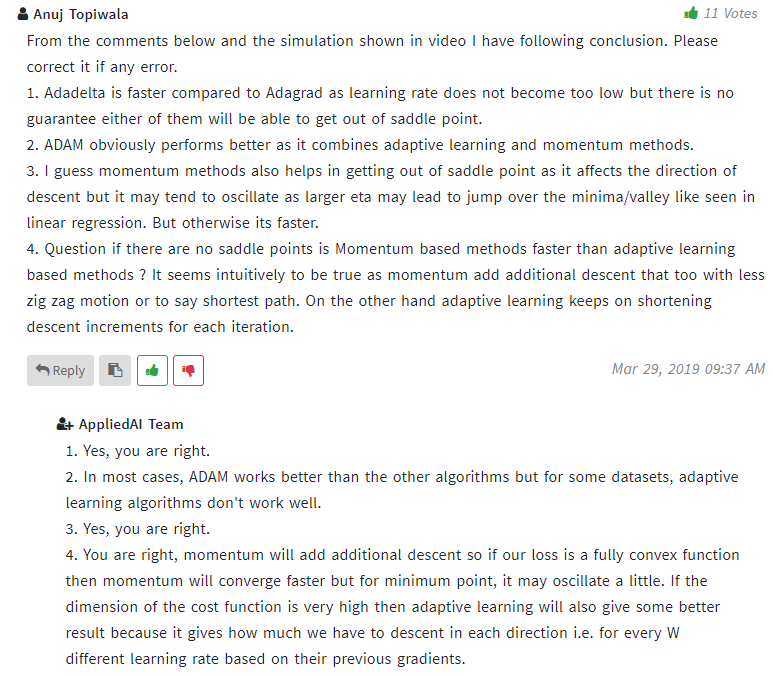
**Mini Batch – SGD**: if NN is shallow and for small dataset.

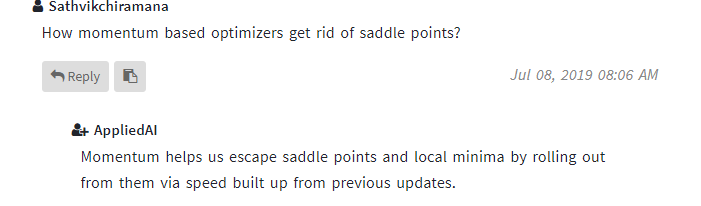
**SGD with momentum or NAG:** It works well for most of the cases but it has comparatively slower convergence in case of non convex function.

**ADAGRAD:** It works well for sparse data(like Computer Vision or Natural language data), and its fast also

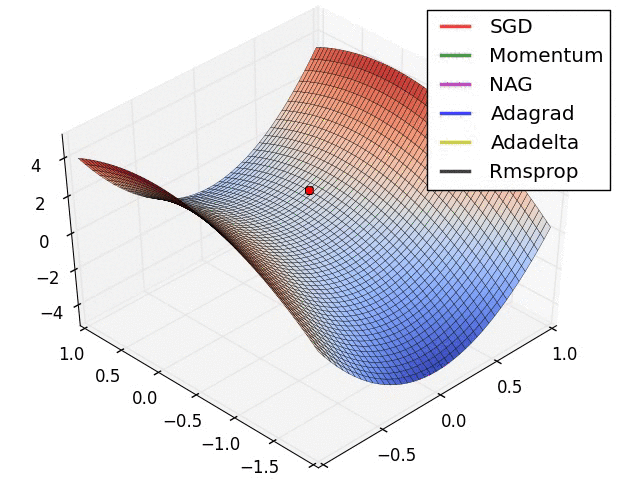
**ADADELTA:** It also works for things ADAGRAD works, along with it’s faster than ADAGRAD as since in ADAGRAD if alpha increase, learning rate will decrease and therefore more no. of iterations will be required.

**ADAM:**  It is best among all as it’s combine the advantage of ADADelta and momentum, and it’s fastest too.





Open it’s html file to see this image in action.



<https://shaoanlu.wordpress.com/2017/05/29/sgd-all-which-one-is-the-best-optimizer-dogs-vs-cats-toy-experiment/>