
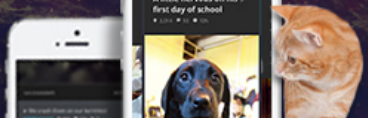


Search  sign in sign up

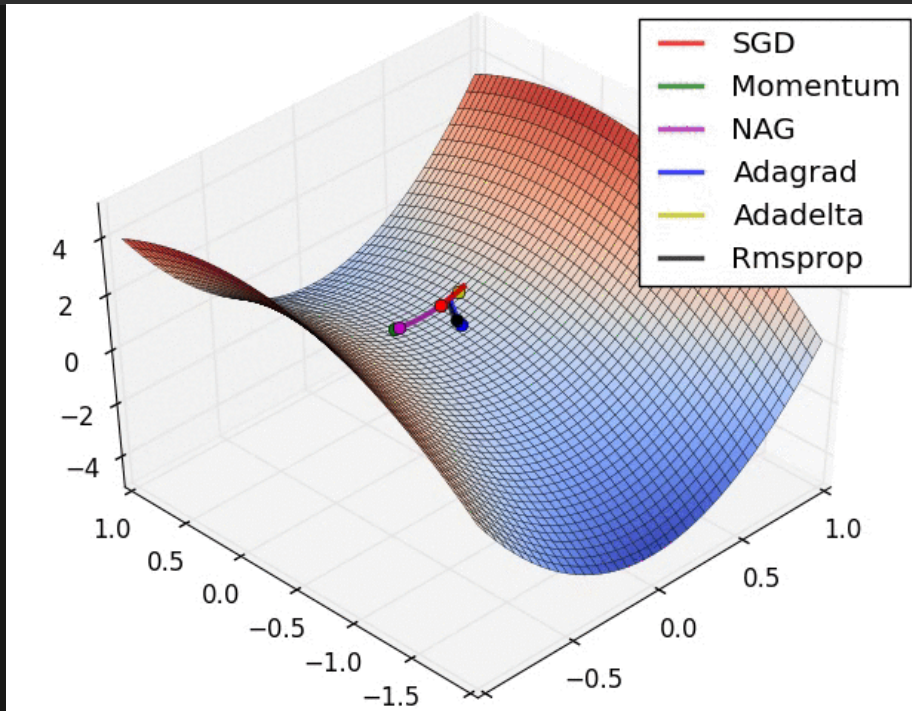


Get the Imgur Mobile App!

Fully Native. Totally Awesome.

[Count me in!](#)

Long Valley



Algos without scaling based on gradient information really struggle to break symmetry here - SGD gets nowhere and Nesterov Accelerated Gradient / Momentum exhibits oscillations until they build up velocity in the optimization direction.

Algos that scale step size based on the gradient quickly break symmetry and begin descent.

Beale's function

Most Viral Images

sorted by popularity



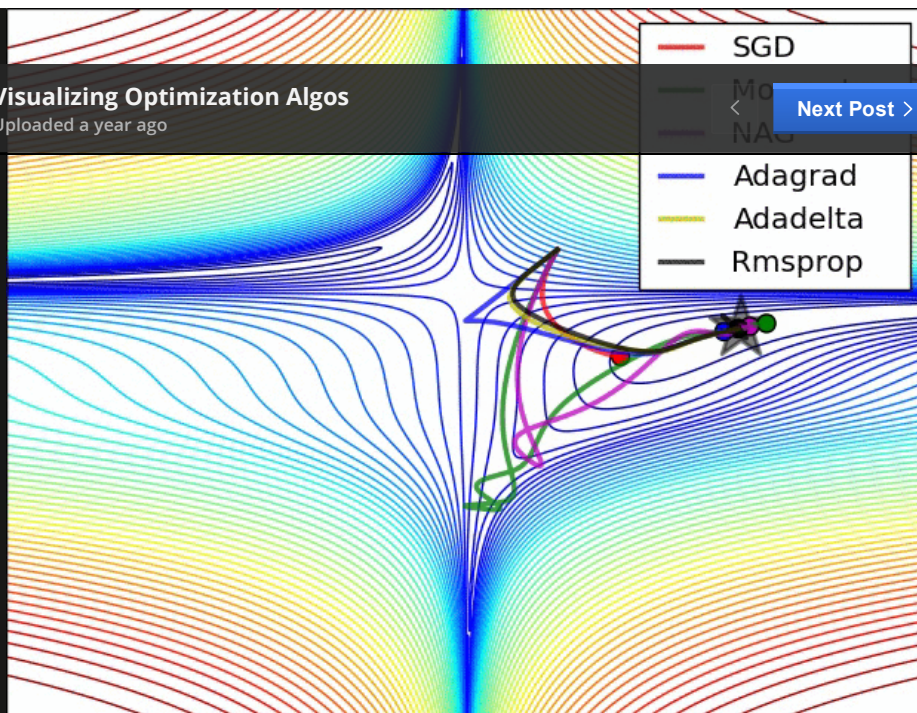
Love Imgur? [Join our team!](#)

[about](#)[store](#)[help](#)[blog](#) *new post!*[terms](#)[privacy](#)[apps](#)[api](#)[advertise](#)[ad choices](#)[request deletion](#)[Follow](#)[Like](#)

Visualizing Optimization Algos

Uploaded a year ago

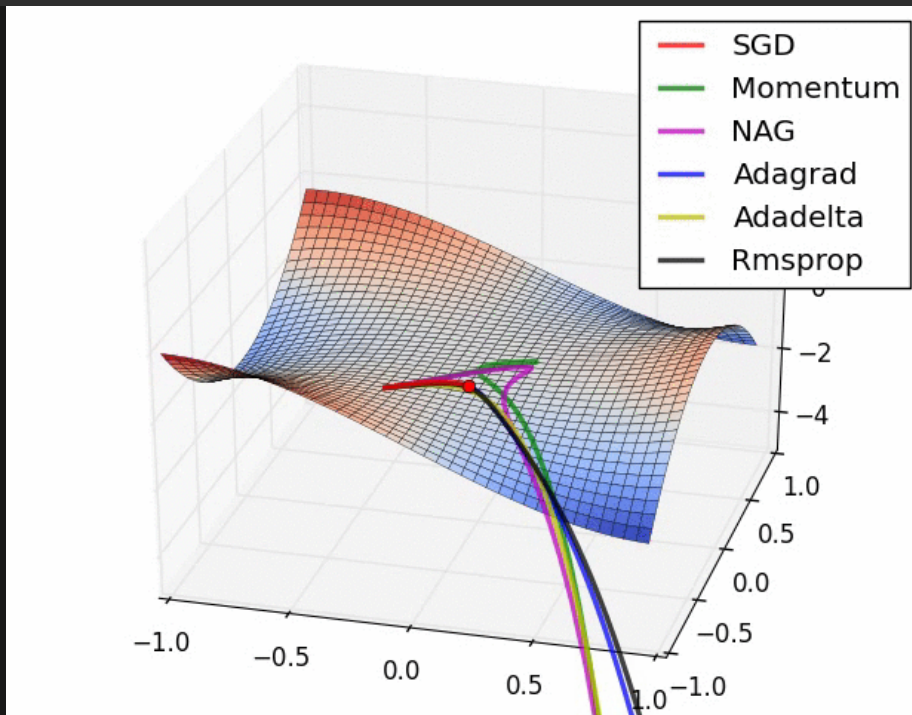
Next Post >



Due to the large initial gradient, velocity based techniques shoot off and bounce around - adagrad almost goes unstable for the same reason.

Algos that scale gradients/step sizes like adadelata and RMSProp proceed more like accelerated SGD and handle large gradients with more stability.

Saddle Point



Behavior around a saddle point.

NAG/Momentum again like to explore around, almost taking a different path.

Adadelata/Adagrad/RMSProp proceed like accelerated SGD.

