Dependency Parsing

Shubham Gupta

January 19, 2020

1 ML and Neural Networks

1.1 ADAM

- Since momentum accumates the gradient from the previous steps, it can help the algorithm get out **pathological curvature** areas. These are areas where the gradient decerasses slowly because the gradient keeps bouncing around the edges of this area, thereby leading to slower convergence. The momentum parameter m helps add momentum from the previous steps and do an exponential average. This is useful because it can use this momentum to go through these curvatures faster, thereby leading to convegence at the global minima.