

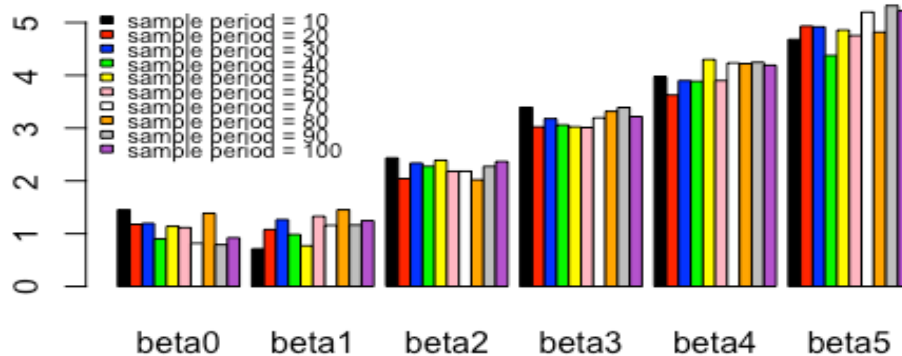
Better Stochastic Gradient Descent

SGD using mini batch

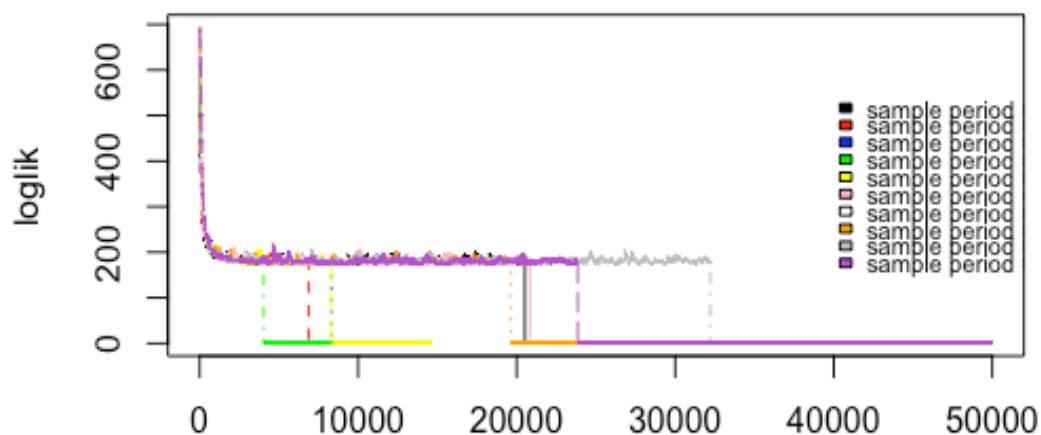
The algorithm updates learning rate (i.e., alpha) periodically based on a small sample from original data. Two parameters are of great interest in this algorithm: (1) sample period specifies how often the algorithms updates learning rate; (2) batch size specifies how much data we sample from original data. We present performance of the algorithm with respect to these two parameters.

Performance w.r.t sample period

coefficient estimation for SGD mini batch

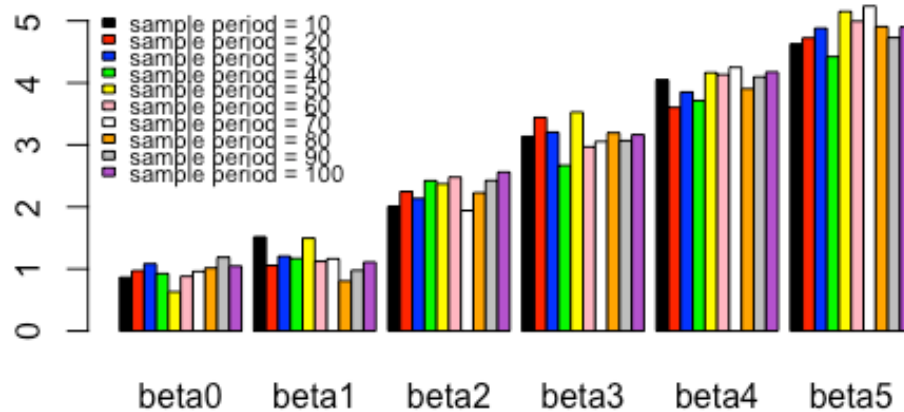


convergence of SGD mini batch w.r.t sample period

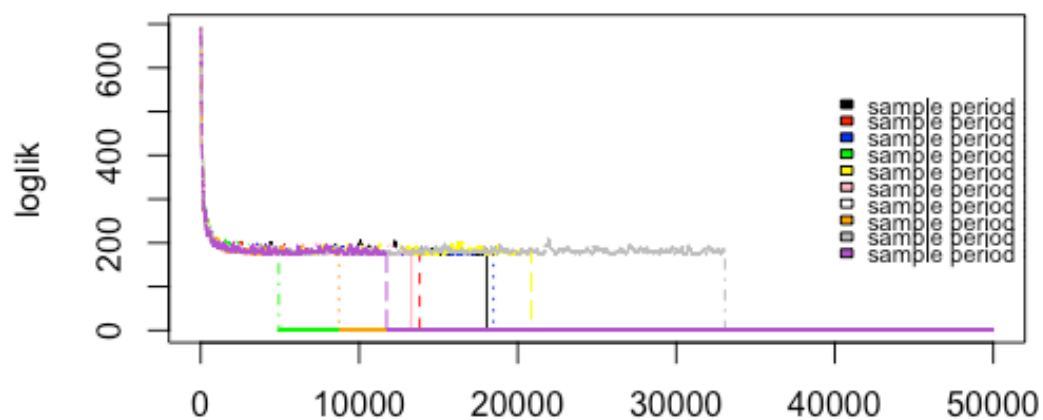


Performance w.r.t batch size

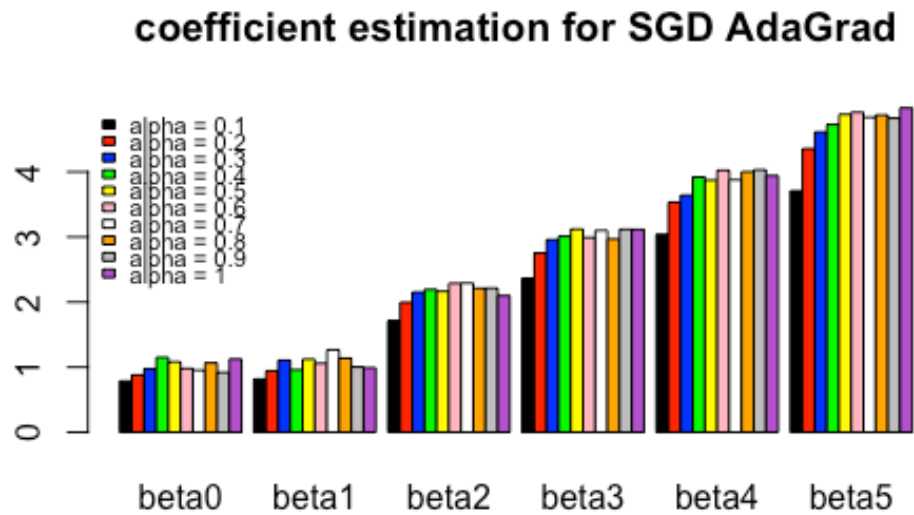
coefficient estimation for SGD mini batch



convergence of SGD mini batch w.r.t batch size



AdaGrad sums previous gradient and takes square root to update new gradient. The parameter of interest is learning rate, alpha.



convergence of SGD AdaGrad w.r.t learning rate

