ADAM : A Method for Stochastic Optimization in ICLR 2015 By Diederik P. Kingma Jimmy Lei Ba

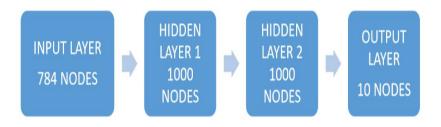
Anshika Chaurasia Razat Shanker

EE18MTECH11017 EE18MTECH11016

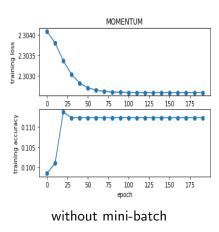
14 Mar 2019

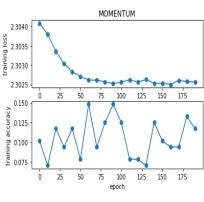
Architecture

ARCHITECTURE OF NEURAL NETWORK



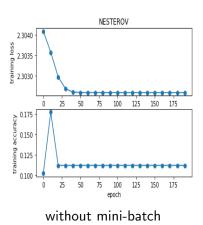
SGD with Momentum





with mini-batch

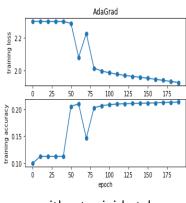
Nesterov



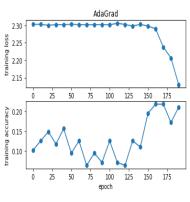
NESTEROV 2.3040 training loss 2.3035 2.3030 2.3025 25 50 75 100 125 150 training accuracy 25 50 75 100 125 150 175 epoch

with mini-batch

ADAGrad

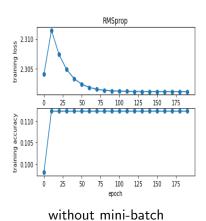


without mini-batch



with mini-batch

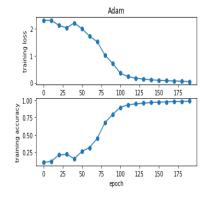
RMSProp



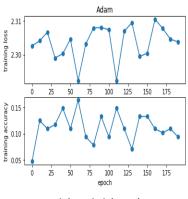
RMSprop 2.30 training 5.30 25 50 100 125 150 training accuracy 25 50 75 100 125 150 175 epoch

with mini-batch

ADAM

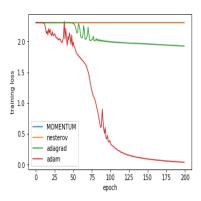


without mini-batch

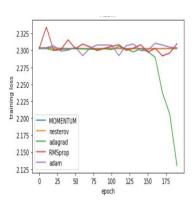


with mini-batch

Comparison



without mini-batch



with mini-batch