

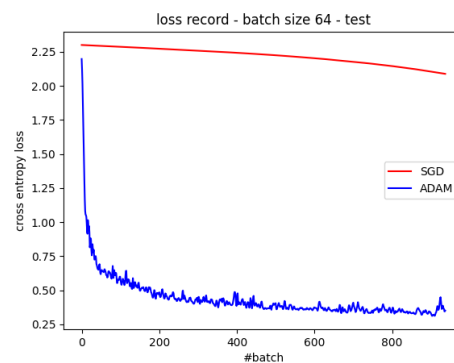
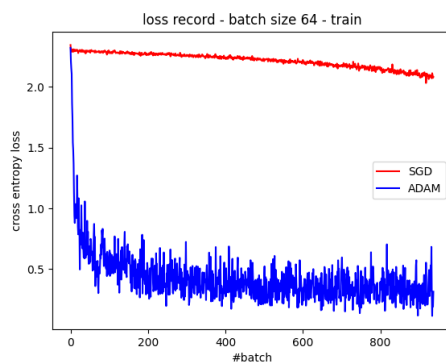
Foundations of Deep Networks

Caspar Schneider, Sinan Kuscü, Sinan Harms

September 2022

PART II - PRACTICE

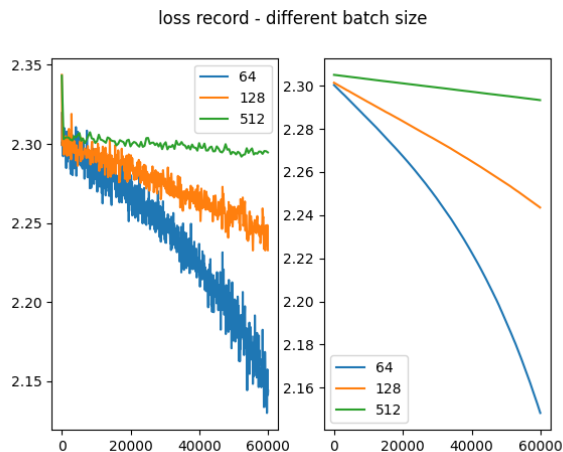
Task1



- SGD slower learning due to randomness of descent
- ADAM faster learning

→ use of both RMSprop and Momentum gradient descent optimization algorithms accelerates training

Task 2



- larger batch size degrades quality of the model
- model is unable to generalize well on data
- smaller batch size improves performance with downside of higher computational cost