

PFN intern 2019 task reports

課題 3、4

SGD, MomentumSGD, Adam を実装した。それぞれの最適化法について、train data、validation data に対する epoch ごとの accuracy 及び epoch ごとの平均損失、batch ごとの損失を記録した。ハイパーパラメーターには課題説明で例示されていたパラメーターを用いた。Adam のパラメーターは基本的に論文に準拠した。

ただし、 $\epsilon=1e-7$ 、 $1e-8$ の 2 つのハイパーパラメーターを使用した。

それぞれの最適化法について、最も高い validation/train accuracy と epoch ごとの平均損失の最終値を表 1 に示した。表 1 と図 1 から MomentumSGD が最も優れた性能を validation accuracy と train accuracy で示していることがわかった。最も Adam の性能が優れていることを予想していたが、 $h=1e-8$ では最終的な accuracy の面では SGD と大差なかった。SGD と比べると Adam のほうが早期に高い validation に対する accuracy を示した。しかし、そこから accuracy が上昇していないことから局所的最適解にはまってしまっていることが考えられる。そこで、Adam に関するハイパーパラメーターを調整することでより性能が上がることを期待した。 $h=1e-7$ を使用した Adam では SGD よりはかなり良い accuracy を示した。

また、loss の初期値を見ると $h=1e-8$ の Adam が圧倒的に高く最初の初期値の問題で収束が遅れた可能性がある。最終的な loss は全ての最適化法の中で validation accuracy の大きさと反比例して MomentumSGD で最も低く、Adam($h=1e-8$)が最も高い値を示した。

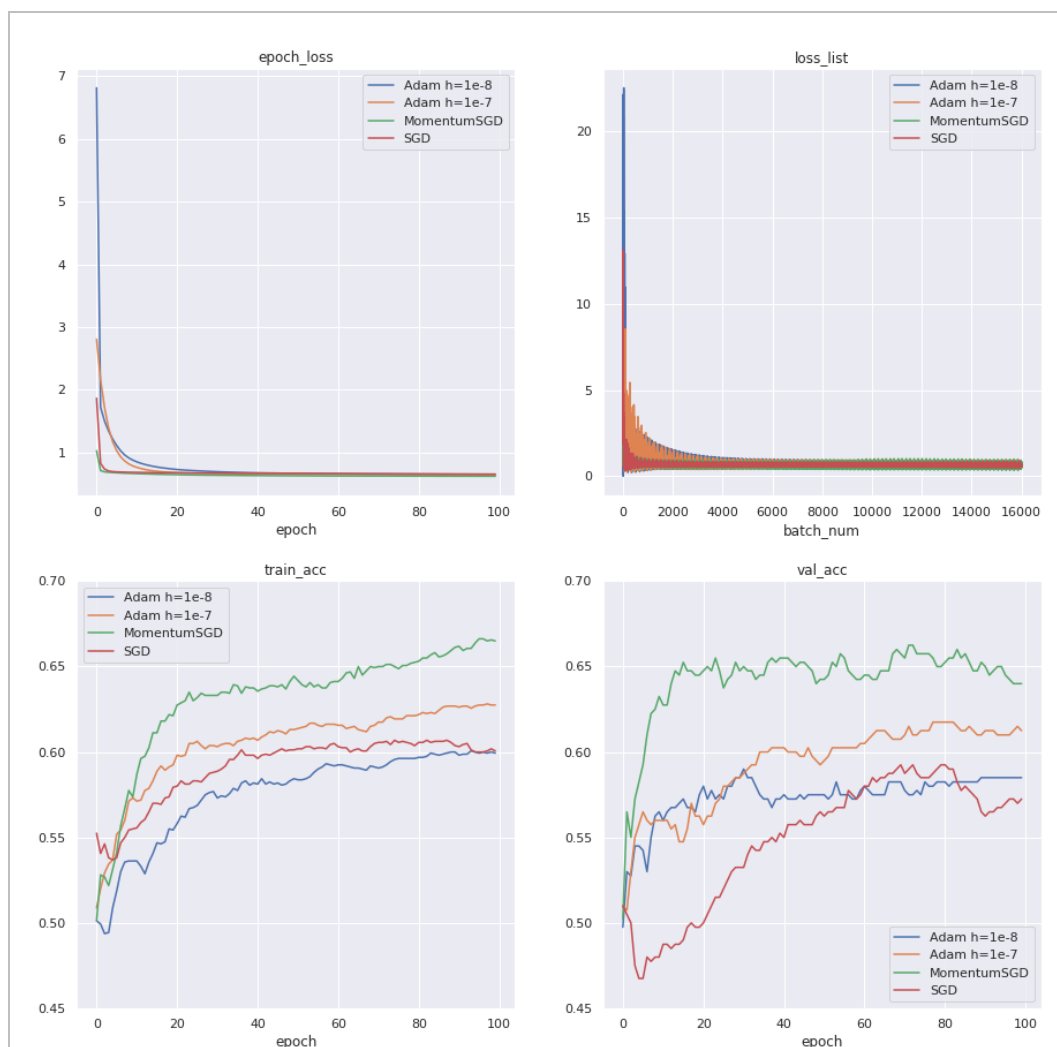


図 1. 最適化法ごとの loss、accuracy のまとめ

表 1. 最適化法ごとの best validation accuracy/best train accuracy/last epoch loss

	Best validation accuracy	Best train accuracy	Last epoch loss
Adam(h=1e-8)	0.59	0.600625	0.64911556
Adam(h=1-7)	0.6175	0.628125	0.64008557
MomentumSGD	0.6625	0.66625	0.61909229
SGD	0.5925	0.606875	0.6521688