

Preferred Networks インターン課題3, 4

Taichi Nishimura

May 2019

1 課題3

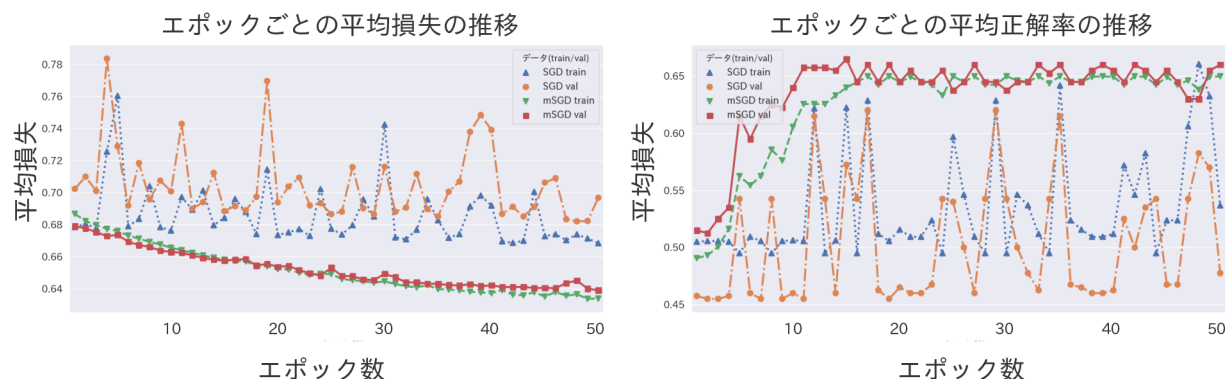


Figure 1: SGD, MomentumSGD での訓練用データ, 検証用データでの性能比較.

課題3では, SGD と MomentumSGD を実装し, 与えられたデータ集合を用いて学習し評価する. まず, 与えられたデータ集合 $D = \{(G_i, y_i) | 0 < i \leq N\}$ (ただし, N はデータ数) を訓練データ D_{train} と検証データ D_{val} に分割した. 実験では, データ集合の8割をランダムに D_{train} に, 2割を D_{val} に分割した. D_{train} を用いてニューラルネットワークを学習させ, 各エポックごとの訓練データでの平均損失, 平均正解率, 及び検証用データでの平均損失, 平均正解率を測定した. なお, エポック数は50エポックとした.

Figure 1 にその結果を示す. 本実験においてはSGD よりも momentumSGD のほうが早く安定して収束していることがわかる. 実験の中で最も性能が良かったのは momentumSGD での15エポック目であり, 検証用データでの平均正解率は66.5%であった.

ニューラルネットワークのハイパーパラメータは以下の通りである. 学習率 $\alpha := 0.001$, 特徴ベクトルの次元数 $D := 8$, モーメント $\eta := 0.9$, バッチサイズ $B := 30$, 集約数 $T := 2$, 数値微分の摂動 $\epsilon = 0.001$ とした. ニューラルネットワークの活性化関数には ReLU 関数を利用している.

2 課題 4

	train loss	train acc	val loss	val acc
SGD+1NN	0.6879	0.6287	0.6888	0.62
mSGD+1NN	0.6581	0.64	0.6574	0.665
Adam+1NN	0.6499	0.6462	0.6476	0.6525
SGD+2NN	0.6734	0.6243	0.6784	0.6025
mSGD+2NN	0.6632	0.6331	0.6556	0.695
Adam+2NN	0.6468	0.64	0.6248	0.7

Table 1: 最適化手法とニューラルネットの深さの変更による性能比較．1NN は課題文中の W が 1 層のニューラルネットワークを，2NN は 2 層のニューラルネットワークを表す．最適化手法は SGD，momentumSGD(mSGD)，Adam を検討した．

課題 4 の中から，最適化手法 (Adam) を変更した場合の性能の変化と，層の深さによる性能の変化を比較する実験を選択した．Table 1 にその結果を示す．この結果より，性能が最も高かったのは Adam+2NN であることが分かった．よって，課題 4 のテストデータでの予想には Adam+2NN の予測ラベルをファイルで保存した (prediction.txt)．ハイパーパラメータは課題 3 と同じものを利用している．また，Adam の β_1 ， β_2 はそれぞれ 0.9，0.999 とし，2NN のネットワークの活性化関数は 2 層とも ReLU 関数を用いて学習した．

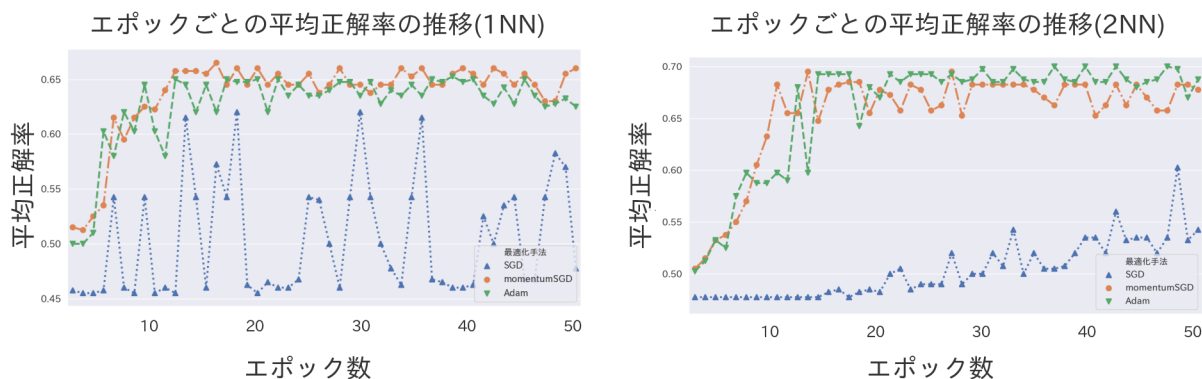


Figure 2: 層の深さと最適化手法を変えた時の検証用データでの正解率の変化

上記の比較手法のエポックごとの正解率の変化を Figure 2 に示す．この結果によると，層を深くすることによって，mSGD，Adam を最適化手法として選択した場合は性能が向上することが分かった．これは，層を深くすることによってモデルの表現できる範囲が高まり，正解率を上げることができたためであると考えられる．また，収束への安定さと速さに着目すると，Adam を利用することでより少ないエポックで安定して収束していることが分かった．