

ニューラルネットワークにおける適応的二次最適化手法

本川 哲哉

近年、層の深いニューラルネットワークを学習させる深層学習は様々な機械学習タスクにおいて大きな成功を収めている。現在、ニューラルネットワークにおける重みパラメータ最適化では、確率的勾配降下法 (SGD) をはじめとする一次最適化手法を用いることが一般的である。しかしながら、一次最適化には収束速度の遅さや、学習率のような調整の難しいハイパーパラメータが多いというデメリットが存在する。これらは主に、一次最適化が損失関数の一次テイラー展開に基づくことに起因する。一次最適化で計算される勾配はノイズや誤差が大きくなってしまふのである。近年この問題を解決するために、損失関数の重みに関するヘッセ行列に基づく二次最適化によって重みパラメータの最適化を改善する試みが活発化している。本研究では Hessian-free optimization (HF) と呼ばれる二次最適化手法を用いて、一次最適化で取り入れられている適応的な工夫を施すことを試みる。一次最適化における適応的な工夫とは過去ステップにおける勾配を蓄積しておき、それらの重み付き平均を用いてパラメータ更新を行うというものである。適応的一次最適化手法にもいくつか種類が存在するが、本研究では Momentum 法と Adam という手法の計算的工夫を適用した。実験には MNIST, CIFAR-10, CIFAR-100 という 3 つのデータセットを用いて画像のクラス分類を行った。また、比較する最適化手法として SGD, Momentum SGD, Adam, HF, HF+Momentum, HF+Adam という 6 種類の手法を選択した。実験の結果、どのデータセットにおいても二次最適化手法の収束の方が一次最適化手法よりも速いことが示された。二次最適化手法どうしを比較すると、HF+Momentum では通常の HF を改善することが難しかったが、HF+Adam では HF を改善することができた。特に CIFAR-10 データセットで HF+Adam が少ない繰り返し数で急速な収束をしており、学習の早期停止と組み合わせることで効率的な学習が行えるようになると考察される。

本研究の実験で使用したニューラルネットワークモデルは、非常に単純な全結合層のみからなるモデルであるため、汎化性能は見込めない。今後の課題としては CNN や RNN のような、より複雑なモデルに対しても HF や Adaptive HF を適用できるようにすることが挙げられる。また本研究では HF のパラメータ更新部分で降下方向の履歴を用いたが、ヘッセ行列を求める部分で過去のヘッセ行列を用いて適応的な工夫を取り入れるということも考えられる。

(指導教員 手塚太郎)