

Preferred Networks インターン選考 2019 コーディング課題 機械学習・数理分野 レポート (課題 3, 4)

佐藤 瞭

課題4では、Adamの実装を選択した。わかりやすくまとめるために、課題3, 4の内容をまとめて、

1. 学習用データ、検定用データそれぞれの平均損失や平均精度の推移
 2. アルゴリズムごと (SGD, momentumSGD, Adam) の平均損失や平均精度の推移
 3. 総括・結論
- の順に解答する。

1. 学習用データ、検定用データそれぞれの平均損失や平均精度の推移の比較

1-1. 平均損失

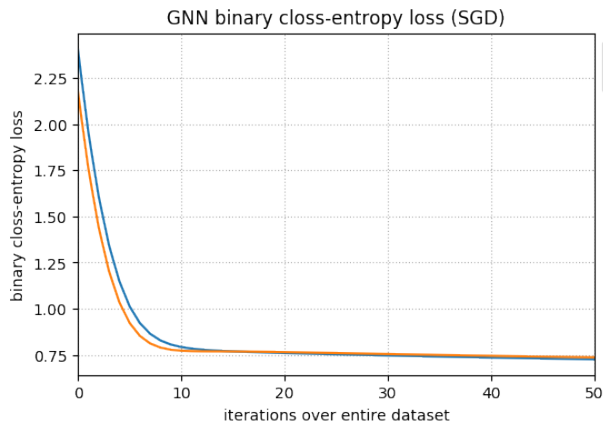


図1 SGDの平均損失の推移

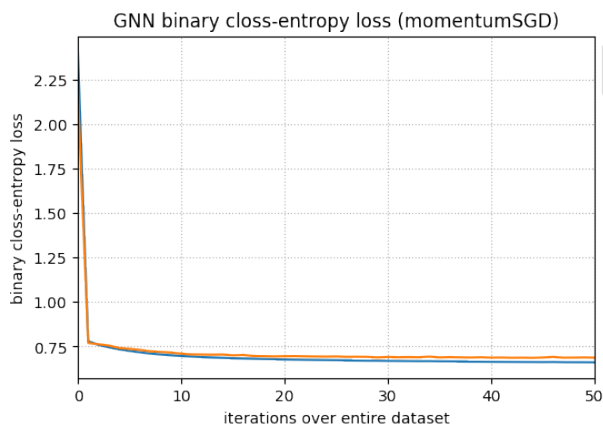


図2 momentum SGDの平均損失の推移

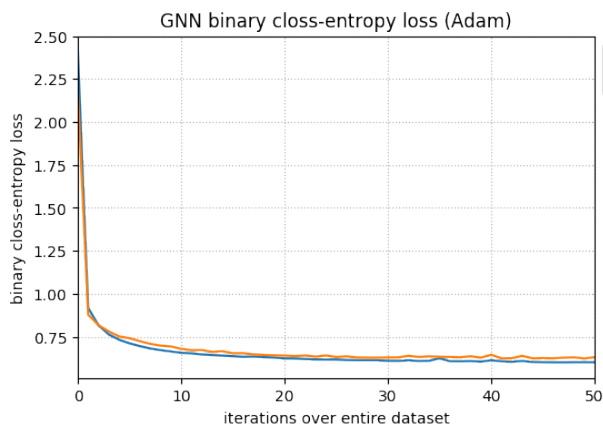


図3 Adamの平均損失の推移

どのアルゴリズムでも、学習用データでの平均損失がほぼ単調に減少していることが確認できる。また、最初の数回の反復では損失の減少の勢いが激しいが、その後は損失の

減少はゆるやかになることも確認できる。さらに、反復を重ねると学習用データの平均損失は検定用データの平均損失よりも小さくなることも確認できる。

1-2. 平均精度

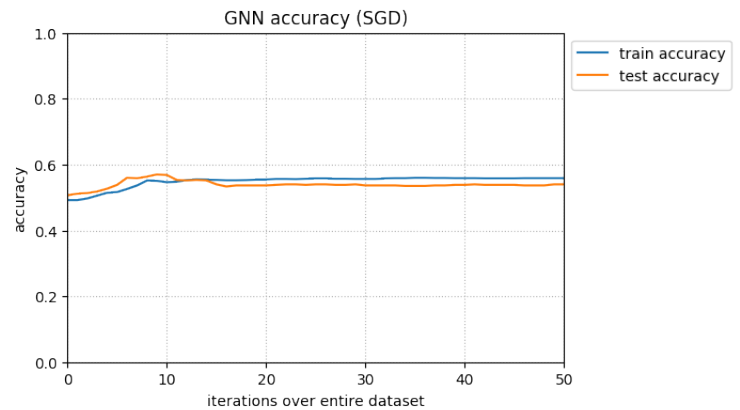


図4 SGDの平均精度の推移

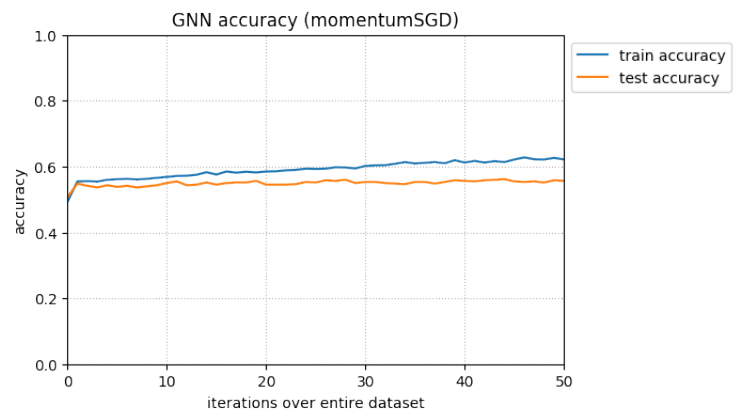


図5 momentum SGDの平均精度の推移

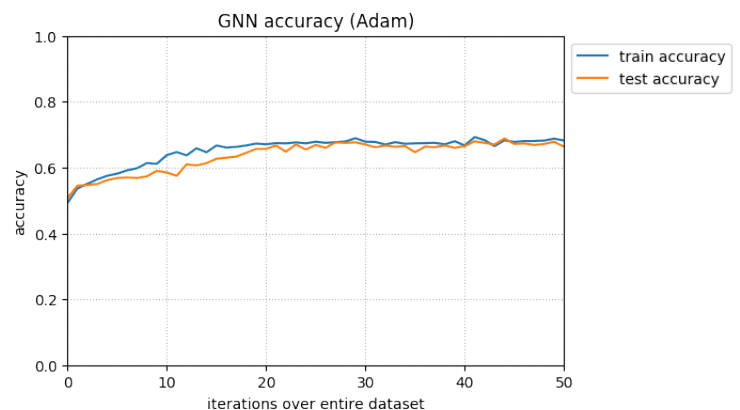


図6 Adamの平均精度の推移

図4を見ると、SGDの平均精度は学習用・検定用どちらのデータでも6割に満たない。反復を重ねても、最初の数回で精度の向上が見られたのみで、それ以降の向上はみられない。つぎに、図5を見ると、momentum SGDの学習用データの平均精度は反復を重ねるごとに向上していくが、検定用データでの平均精度はあまり向上せず、反復を重ねるごとに過学習ぎみになっていくことが伺える。図6のAdamの結果を見ると、学習用・検定用の両方のデータの平均精度は反復を重ねるごとに6割〜7割のあたりまでは向上していることがわかる。また、学習用・検定用のデータの平均精度の乖離はみられないため、Adamによるモデルは汎化性能が高いと考えられる。

2. アルゴリズムごと (SGD, momentumSGD, Adam) の平均損失や平均精度の推移の比較

2-1. 平均損失

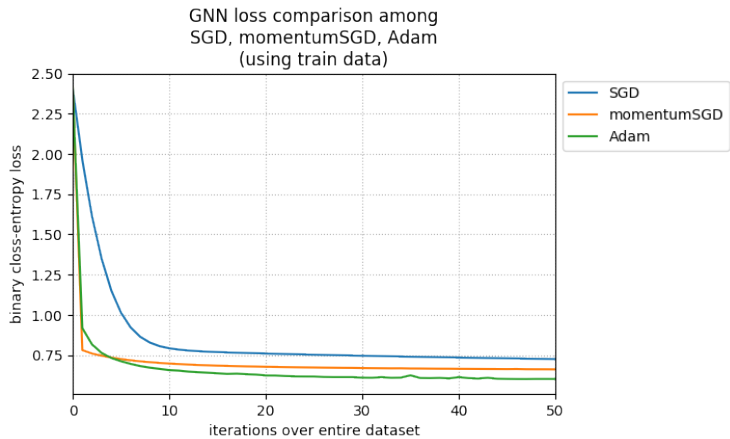


図 7 学習用データでの平均損失の推移

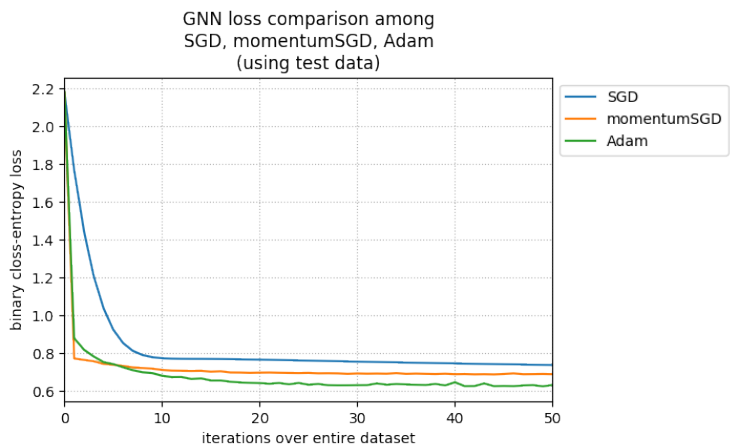


図 8 検定用データでの平均損失の推移

図 7・図 8 をみると、反復回数を大きくしていくと Adam < momentum SGD < SGD の順に損失が小さくなることがわかる。ただし、反復を始めて何回かの間は、momentum SGD の損失が勢い良く減少して Adam の損失を下回っている。また、SGD の損失の減少の勢いは他のアルゴリズムと比較して小さいこともわかる。

2-2. 平均精度

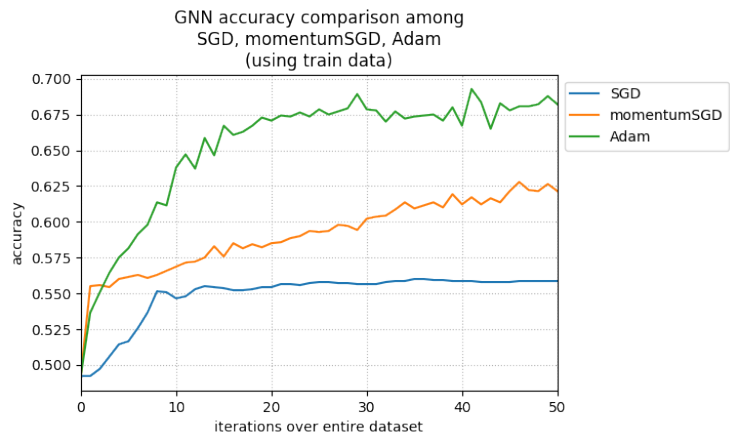


図 9 学習用データでの平均精度の推移

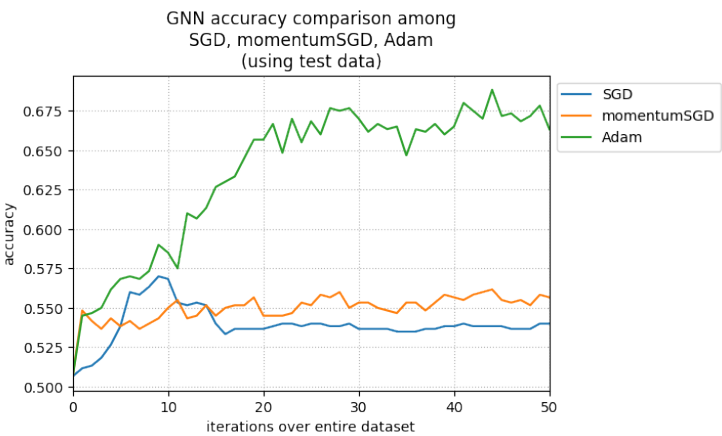


図 10 検定用データでの平均精度の推移

図 9・図 10 をみると、反復回数を大きくしていくと Adam > momentum SGD > SGD の順に精度が良くなることがわかる。ただし、反復を始めて何回かの間は、momentum SGD 学習用データが勢い良く増加して Adam の精度を上回っている。しかし、検定用データの momentum SGD と Adam の、反復回数が少ないところでの精度には学習用データでみられたほどの差がみられないため、これは momentum SGD の過学習によるものである可能性が考えられる。つぎに、図 9・図 10 を比較し、学習用データの精度と検定用データの精度の差に注目してみる。SGD や momentum SGD は、反復回数を大きくしていくと精度の差が大きくなっていくことがわかる。そのため、この 2 つによるモデルは、反復回数を大きくするほど過学習していくと考えられる。一方、Adam の精度に注目すると、反復が 5 回～20 回のあたりで、検定用データの精度が学習用データの精度を明らかに下回っている。しかし、反復が 20 回以降では両者の差は小さくなっている。これは、反復により過学習が解消されていくということだろうか。

3. 総括・結論

以上の結果を踏まえると、今回実装した SGD, momentum SGD, Adam の中で、精度や汎化性能に優れるのは Adam である、と結論づけることができる。ただし、学習回数を非常に少なく (5 回以下程度) しなければならないような制約がある場合は、反復初期での精度向上に優れている momentum SGD も有用かもしれない。