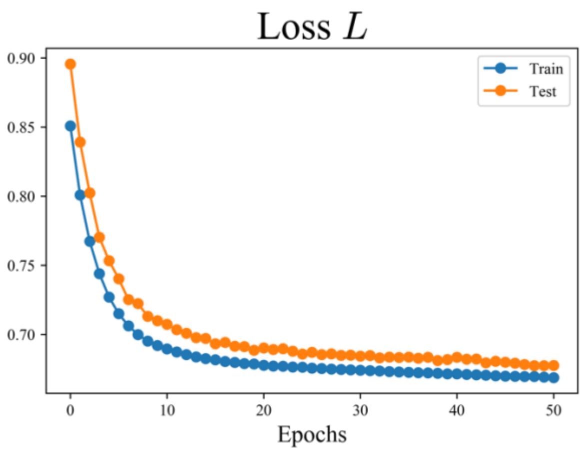
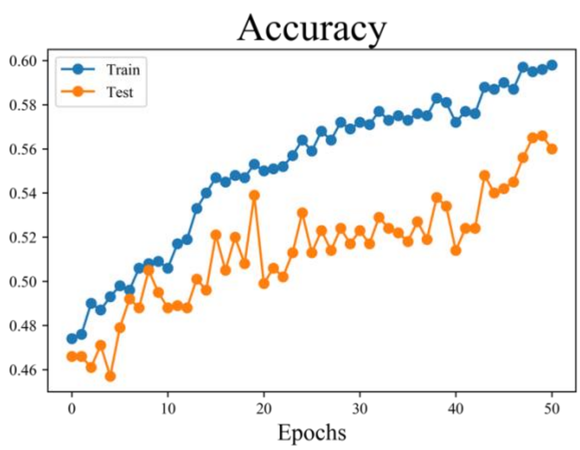
・**課題 3**

計算設定 ： 学習率$\epsilon = 0.001$、微分微小量$\delta = 0.001$とした。Momentum-SGD の慣性項係数には$\eta = 0.9$を用いた。ハイパーパラメータW, Aは正規分布からランダムサンプリングし SGD, Momentum-SGD で同一の初期値を用いた。2000 個のデータを学習データとテストデータの半分ずつに分け、ミニバッチサイズを 10 として 50 エポックの学習を行った。

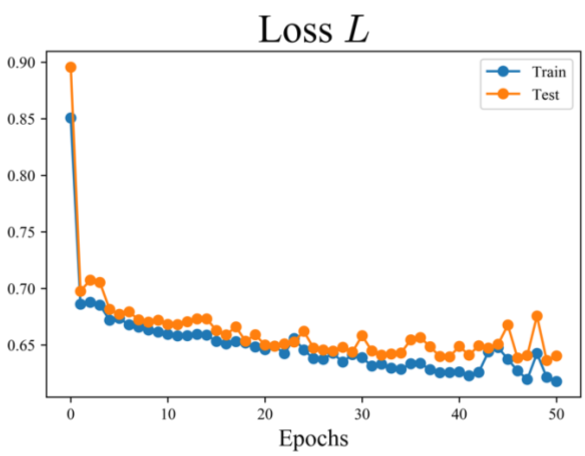
結果 ： 学習過程の平均損失(Loss)、正答率(Accuracy)はそれぞれ次のグラフのようになった。

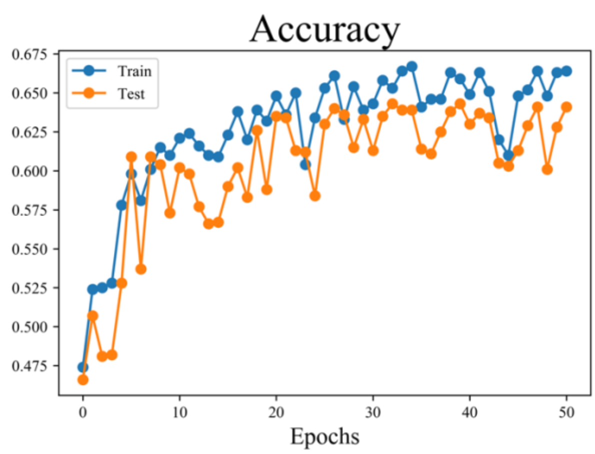
- SGD





- Momentum-SGD





平均損失について、学習初期に Momentum-SGD は SGD と比べて急速に減少している。これは慣性項によりパラメータ更新量を大幅に増加させたことに起因する。ただし、慣性項の影響により Momentum-SGD では収束段階に入った 40epochs~以上で不安定に上下する振る舞いが見られ、逆に SGD では安定した収束が実現している。平均損失と同様に、正答率に関しても Momentum-SGD では急速な上昇と不安定な収束が見られる。SGD は正答率が 50epochs においても収束していない。全体的に、損失・正答率ともに Momentum-SGD が高パフォーマンスを示している。

これら 2 手法の比較から、Momentum-SGD において慣性項や学習率を徐々に減少させていくような操作を加えることで学習初期における急速な損失関数の減少と学習後期の安定した収束を共に実現するような学習アルゴリズムにすることができると予想される。

・**課題 4**

計算設定 ：GNN のパラメータ更新に Adam アルゴリズムを用いたものを実装した。SGD ではハイパーパラメータθ_tを、損失関数微分$g_t \equiv \nabla_\theta L_t(\theta)$として$\theta_t \leftarrow \theta_{t-1} - \alpha g_t$のように更新していた。Adam では、

$$\theta_t \leftarrow \theta_{t-1} - \alpha\, E[g]/E[g^2]$$

のようにgの期待値を用いて更新する。期待値は初期データの重みが指数的に減少するような移動指数平均によって求める。具体的な更新アルゴリズムは以下のように実装した。

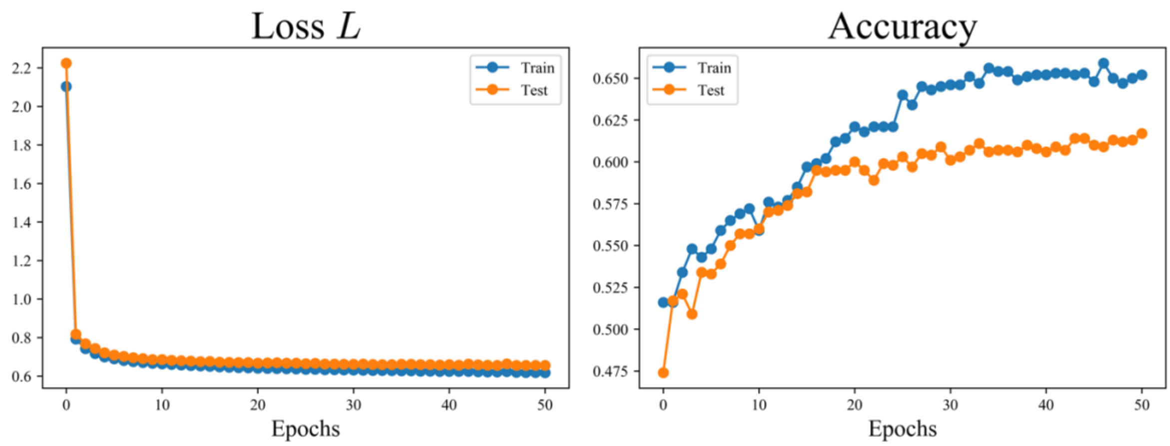
$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2$$

$$\theta_t \leftarrow \theta_{t-1} - \alpha \frac{m_t}{1-\beta_1^{t'}} \sqrt{\frac{1-\beta_2^{t'}}{v_t + 10^{-9}}}$$

ここで、t'は epoch 数を用いた。また、パラメータとして$\epsilon = 0.001, \alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.99$を用いた。

結果 ：2000 個のデータセットを訓練用 1000,テスト用 1000 に分け、ミニバッチサイズ 10 として 50 エポックの学習を行った。学習過程の平均損失および精度(正答率)は以下の図のようになった。



平均損失は訓練・テストデータ共に単調に減少しており、正答率も増加しているため学習が上手く進んでいることがわかる。およそ 30epochs 程度から正答率の収束が見られ、テストデータに対する精度は60%以上に達している。50epochs 後のハイパーパラメータを用いてラベル無しデータに対する予測計算を行った。