

多層パーセプトロンモデルによる手書きの数字認識

東北大学大学院医学系研究科 博士課程 田中 真一

2024.04.17

2024.04.18 修正

2024.04.23 修正

序

深層学習を理解するための初歩段階として、MNIST¹⁾ を用い、多層パーセプトロンモデルによる手書き数字の認識を行った。損失関数には交差エントロピー誤差関数を、最適化には確率的勾配降下法を用いた。計算には Python 3.11.2²⁾ + PyTorch 1.13³⁾ を使用した。

学習の概略

今回は input layer, 2 層の hidden layer, および output layer から成るモデルを構築した。ノード数は input layer が 784, hidden layer が 1024 と 512, output layer が 10 であり、いずれも全結合層である。学習には MNIST の train データセット (サンプル数 60,000) を用いて、256 サンプルずつのミニバッチ学習を行った。エポック数は 10 とした。

テストデータセットによる評価

作成したモデルを用いて MNIST のテストデータセット (サンプル数 10,000) について数字の推定を行ったところ、正しく認識できたものが 9,019 例、誤って認識したものが 981 例であった。その一部の例を output1, output2, output3 に示す。括弧内はモデルの出力値である。

次に、この MNIST で学習したモデルを用いて、筆者が作成した「6」が書かれた数字画像 30 例について数字の推定を行ったところ、正しく 6 と判定できたのは 6 例のみであった。これらの結果を output4, output5, output6 に示す。

さらに、MNIST のテストデータセットからランダムに選んだサンプルについて、上下左右に数字の位置を平行移動してから数字の推定を行った。この場合、移動量が小さい場合には概ね正しく判定できたものの、移動量が多い場合には誤判定が多くなった。これらの結果を output7, output8, output9 に示す。

考察

MNIST のテストデータセットについて、正答率が 90% であった。簡易なモデルではあるが、ある程度は正しく推定できており、モデルの構築自体はよくできたものと考ええる。

筆者が作成した画像データに対しては正答率が 20% と、著しく低かった。これは MNIST に収録されてい

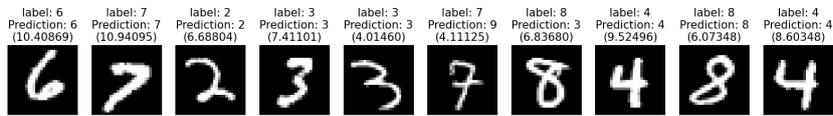


図1 output1

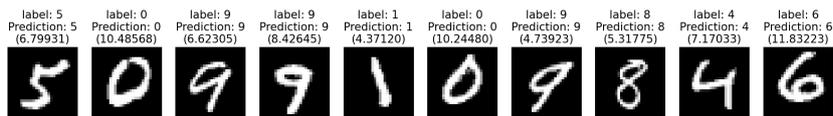


図2 output2

る数字とは、意図的に大きく書体が異なる数字を用いたためと考えられる。深層学習、特に今回の多層パーセプトロンモデルのような簡易なモデルにおいては、学習データと似ていないデータに対しては適切な推定を行えないことを確認できた。

興味深いのは MNIST のテストデータを平行移動した場合の結果である。人間の眼では、平行移動してもしなくても同じ数字に見えるし、いずれも正しく判読することは容易である。しかし今回構築したモデルでは、平行移動により正しい推定を行うことができなくなった。これは、多層パーセプトロンモデルでは画像に記されている図形の形態を一切解釈しておらず、「どのあたりが白でどのあたりが黒いか」という情報のみを用いて数字の推定を行っているからである。この「図形として解釈していない」という点は、人間による文字認識と大きく異なる。

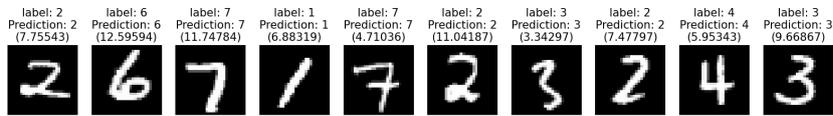


図 3 output3

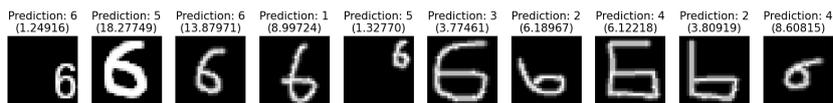


図 4 output4

まとめ

多層パーセプトロンモデルによる手書き数字の推定を行った。よく整えられたデータセットに対しては相応の精度で数字を判定することができた。また、数字を平行移動した画像をテストに用いることで、数字を図形としては認識していないというモデルの弱点を鮮明に示すことができた。

参考文献

- 1) The MNIST Database of Handwritten digits, <http://yann.lecun.com/exdb/mnist/> (2024.04.17 閲覧)
- 2) Python.org, <https://www.python.org/> (2024.04.17 閲覧)
- 3) PyTorch, <https://pytorch.org/> (2024.04.17 閲覧)

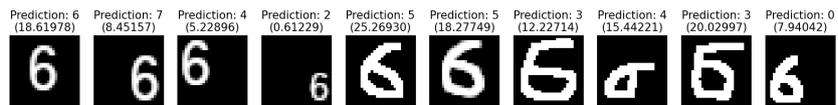


图 5 output5

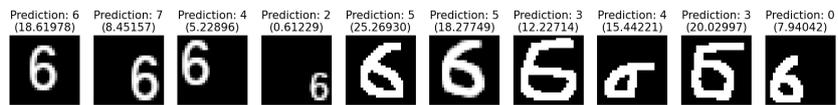


图 6 output6

index: 588
label: 5



7 output7

index: 6755
label: 8



8 output8

index: 8124
label: 7



图 9 output9