COMP 551: Applied Machine Learning Report for Project 3

Zijian Pei, Xiaoxiao Shang, Yueng Zhang

April 4, 2022

Contents

1 Abstract

The main task for this project is to implement the multilayer percepton model in order to classify the image data and find out how different methods of training data can effect the model performance in the Fashion-MNIST database. We try to add hidden layers, train with different activation functions, add dropout regularization(L2) to the network,use unnormalized images and then compare the performance with CNN(convolutional neural network). Here are the most important findings after we conducted the experiments. First, adding hidden layers can increase the accuracy of MLP,whereas more layers doesn't guarantee a better performance. What's more, training model with different activation functions will have little change in performance. Also, having L2 regularization may have some negative effect on the performance, and using unnormalized data will dramatically decrease the performance.Then, training data with CNN has a similar performance compared to MLP. Finally, we found that the best model is achieved by implementing 1 hidden layer which has a test accuracy of 88.77%.

2 Introduction

Sight is the sense which humans rely on the most. It's used in all facets of life from driving to recognizing objects and faces. While this is an intuitive task to a human, image recognition is much more difficult for computers.[1] Using multilayer percepton, we are able to map each clothes to its label. The dataset Fashion-MNIST has sufficient images for training and testing the model.

From our experiments, we noticed that the model with the highest accuracy,88.77%,is achieved by establishing 1 hidden layer MLP with 128 units, batch number of 70 and 0.05 learning rate. Besides, the accuracy using one and two hidden layers are both about 12% higher than that of zero layer, showing that adding layers is one of keys to improve performance. In contrast, the accuracy of using leaky ReLU is 2% higher than using tanh, which indicates the activation function doesn't effect the performance generally. Then, using L2 regularization will drop accuracy when number of iterations are less than 20, and after that the performance are resembling to normal MLP. Last, the experiment of training model with unnormalized model demonstrates a noticeably diminish of the performance (13%), and that shows the significance of normalizing database before training.

3 Datasets

Fashion-MNIST is a dataset of Zalando's article images—consisting of a training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28x28 grayscale image of clothes, associated with a label from 10 classes(0-9).[2]

In this project, we first vectorized each data to a 1-D array with size of 784. Then, we normalized data by first subtracted mean then divided by standard deviation. After that, we got our training and test set with zero-centered. We will see the effect of handling the data this way in the following experiments.

4 Results

4.1 Experiment 1: Different number of hidden layers

From the experiment, we can tell that MLP with 1 or 2 hidden layers will have higher test accuracy than MLP with 0 hidden layer within the same number of iterations. The curves of MLP with 1 and 2 hidden layers are similar, and the test accuracy converge after about 20 iterations. MLP with 0 hidden layer converges much slower than the other two models, and it does not converge after 100 iterations.

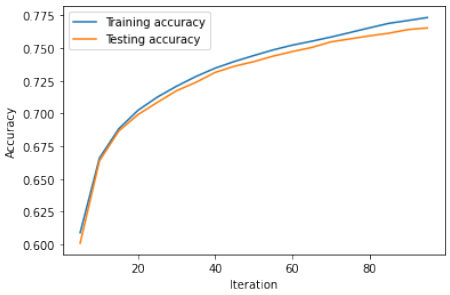

Fig. 1. MLP with 0 layer

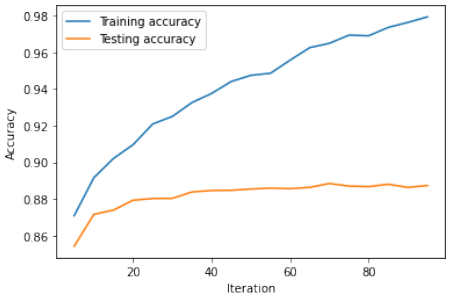

Fig. 2. MLP with 1 layer

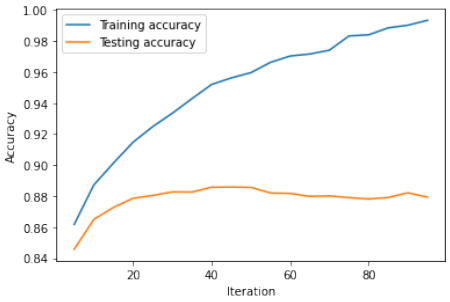

Fig. 3. MLP with 2 layers

4.2 Experiment 2: Different activation functions

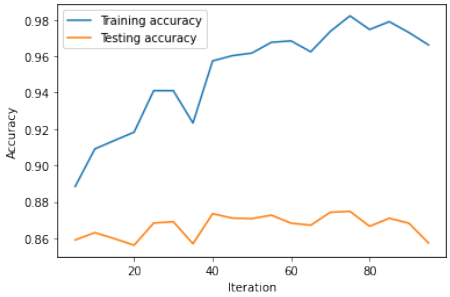

Fig. 4. MLP with tanh

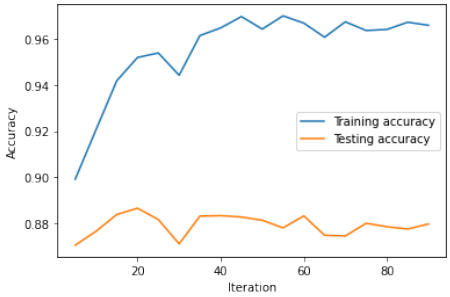

Fig. 5. MLP with Leaky-ReLU

From the experiment, we can tell that MLP with activation function tanh will have slighly lower test accuracy (around 0.86) than MLP with activation functions ReLU and Leaky-ReLU (around 0.88).

4.3 Experiment 3: Add L2 loss to activation function

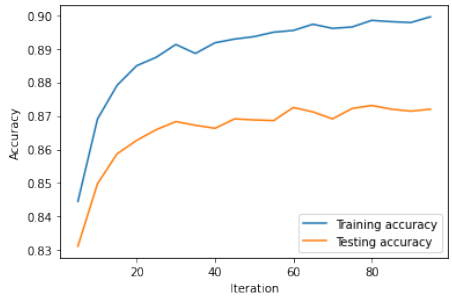

Fig. 6. MLP with L2 loss

From the experiment, we can tell that adding L2 regularization to MLP sometimes may decrease the performance. The accuracy for adding L2 regularization is smaller than MLP with 2 hidden layers, but there is a rapid increase from 0 to 20 iterations. At the end, performances of adding L2 regularization is 0.8731 which is really close to the normal MLP which is 0.881.

4.4 Experiment 4: Train with unnormalized images

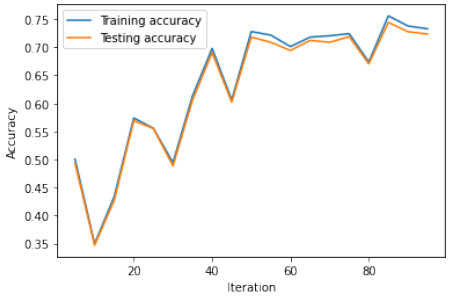

Fig. 7. MLP with unnormalized images

From the experiment, the accuracy of using unnormalized data is 0.7553. we can tell that will drop out the accuracy to a great extent.

4.5 Experiment 5: CNN with 2 convolutional and 2 fully connected layers

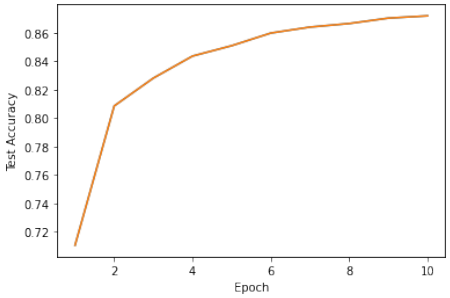

Fig. 8. CNN

From the experiment, the accuracy of using convolutional neural network (CNN) with 2 convolutional and 2 fully connected layers is 0.8718, and compared to experiment 1, the accuracy of MLP with 2 hidden layers is 0.881. We can tell that the performance of using CNN is similar to MLP.

4.6 Experiment 6: MLP with best performance

In this classification problem, MLP with one hidden layer has smaller computation complexity than MLP with two hidden layers; and MLP with 0 hidden layer converges much slower based on the results from the first experiment. Therefore, we decided to implement our models using MLP with 1 hidden layer. The next step is to determine the proper setting of learning rate and batch number. During the process of running the above experiments, we noticed that a learning rate greater than 0.1 typically should cause stack overflow. As a result of that, a value incremented by 0.01 starting from 0.01 is used to measure the most suitable value of learning rate. To determine the proper setting of batch number, we choose the initial batch number 10 and incremented by 10 each time until 100. The experiment shows that the best learning rate is around 0.05, and the batch number equals 70 will generate the best performance which has accuracy of 0.8877.

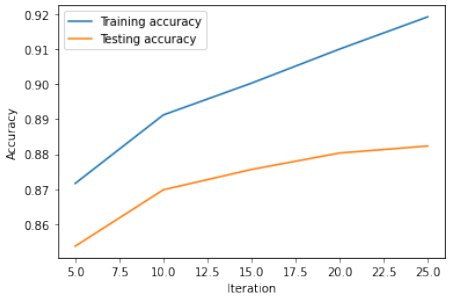

Fig. 9. Learning rate 0.05 and 30 iterations

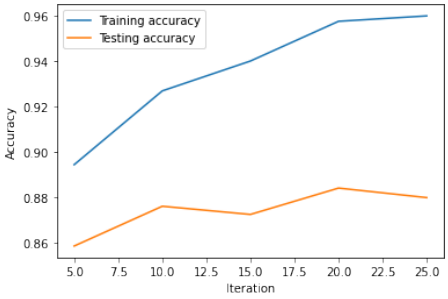

Fig. 10. Batch number 30 with learning rate of 0.05

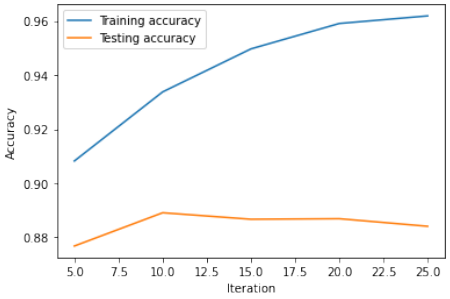

Fig. 11. Batch number 50 with learning rate of 0.05

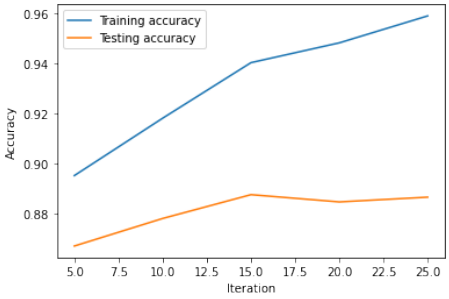

Fig. 12. Batch number 70 with learning rate of 0.05

5 Discussion and Conclusion

From experiment 1, we can conclude that non-linearity can improve the accuracy, but adding more layers will not further promote it. This is not we expected. After discussion, we find that is due to unsophisticated pattern of the Fashion-MNIST. Since 1 layer MLP is already sufficient and has a good performance, there is no need to add more layers.

From experiment 2,3 and 5, we didn't find obvious effect on the performance by changing the activation function,adding regularization and using CNN with 2 convolutional and 2 fully connected layers.

Experiment 4 shows that the accuracy has a huge drop if we don't normalize dataset. According to our knowledge, we discover that the outliers may lead the gradient too far, resulting in NaN values. Therefore, the training process gets worse.

What's more, in experiment 6 we tried to come up with an MLP architecture that performs as good as possible. We construct the model with 1 hidden layer, batch number of 70 and 0.05 learning rate with accuracy of 0.8877.

Finally, if we can investigate the model further, we would try to use other regularization techniques such as L1 regularization and try to implement CNN with pooling.

6 Statement of Contributions

Xiaoxiao Shang: Report writing, Section 1-2 implementation
Yuteng Zhang: Report writing, Section 1-2 implementation
Zijian Pei: Report writing, section 3.1-3.6 implementation

References

[1] Edwards Yen. *Fashion-MNIST*. Website. https://cnedwards.com/files/Edwards_Yen_Final_Project.pdf. 2020.

[2] Edwards Yen. *Fashion-MNIST*. Website. https://github.com/zalandoresearch/fashion-mnist. 2020.