Project 1: Discussion of Neural Network into MNIST

Chengrui Wang 22300680231
Fudan University
22300680231@m.fudan.edu.cn
https://github.com/wcrisdog/PJ1
https://drive.google.com/drive/folders/1Z4mBreBQI9Znx1ALYvuj4ymB1dmF9DFC

Abstract

In this report, I will present the trials and errors of different modules, model structures, learning rate schedulers, optimization methods, regularization methods, and some more advanced skills mentioned in lectures and some previous papers. The github repo and related google drive where the dataset and model are stored are also attached above, below the personal information. Throughout the process, it's interesting to see how everything changes and the model being improved in a sense.

1 Introduction

Neural network could be viewed as a generalized way of fitting functions, mostly in a non-linear way. Throughout history, it has undergone great development, with several times of name change according to the book by Goodfellow [2016].

Also, MNIST, by Deng [2012]), stands for Modified National Institute of Standards and Technology. It consists of scans of handwritten digits and associated labels describing which digit 0–9 is contained in each image. Its popularity and relative simplicity provide great opportunities for beginners like to try out all different methods. Though it does seem that some naive methods behave good enough after some batches of training, it still feels good to practise the different skills, which provides us with a toolkit in lots of other more complicated problems.

2 Preliminary

2.1 Dataset Exploration

To start with, in order to get a glimpse of how the image dataset looks like, I read a chunk of the final pictures to get a sense of the data structure, shape of different variables, and a visualization to boost myself up. The images I visualize are shown in 1:

Afterwards, I also see the shape as well as the content of the labels attached to each picture.

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

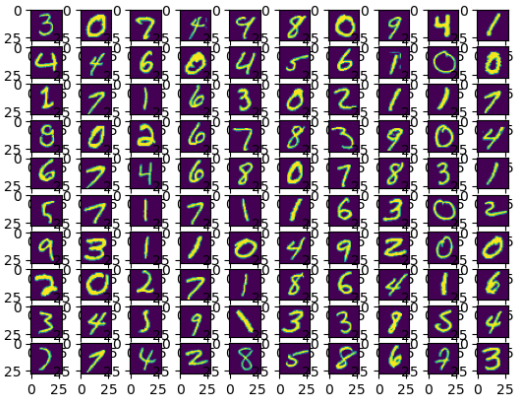

Figure 1: Sample Graph of handwritten digits

2.2 Summary of Tests

1. Model:
   - MLP
   - CNN
2. Learning Rate Schedulers:
   - Step Adjustment
   - Multi-step Adjustment
   - Exponential Adjustment
3. Optimization Methods:
   - SGD
   - MomentSGD
4. Regularization Methods:
   - l_2 Regularization
   - Dropout
   - Early Stopping

3 Implementation Details

The training and testing process involves the following steps:

1. **Load Data**
2. **Shuffle and Split**: Use the first N as a validation set, and the remaining as the training set
3. **Model**: Set up a specific model with the corresponding layer sizes. (For MNIST, it's [784, 600, 10])
4. **Optimization**: With a certain method to adjust the learning rate, create a certain optimizer
5. **Loss and Metric**: Here we use the Cross Entropy Loss with softmax as its final layer
6. **Train and Visualize**: Train the model, see the test result and visualize the weight to pinpoint any problem inside there.

This structured approach ensures robust training and testing, allowing us to fine-tune parameters for optimal performance.

4 Result

For a simple baseline, I use the MLP model, with the default MultiStepLR and milestones given, SGD optimizer with L_2 regularization.

Table 1: Comparison of different experiment setup

Category	Method	Val. Acc. (%)	Test Acc. (%)
Model	MLP	92.3	91.8
	CNN	98.1	97.6
Scheduler	StepLR	94.5	94.0
	MultiStepLR	92.3	91.8
	ExponentialLR	93.7	93.2
Optimizer	SGD	92.3	91.8
	MomentumSGD	96.2	95.8
Regularization	L_2	92.3	91.8
	Dropout	96.5	96.0
	Early Stopping	97.2	96.8

5 Conclusion and Future Works

In this report, with the usage of MNIST dataset, we test the performance of different models with different setups. And this points to similar results due to the simplicity of this problem.

Limitations: Though we try out different methods, we didn't perform data augmentation, some more advanced regularization and optimization methods.

Meanwhile, for some models, I do observe some instability, which might be out of implementation issues.

References

Li Deng. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE signal processing magazine*, 29(6):141–142, 2012.

AC Ian Goodfellow. Deep learning-ian goodfellow, yoshua bengio, aaron courville-google books, 2016.