MNIST Classification with Softmax (score 10)

Homework 1 for Deep Learning, Autumn 2025

Deadline: 2025.10.13 12:00

Attention

- You need to submit **a short report** (at least one page **in PDF format**), and **all codes** of the programming practice. The report should include the answers of objective questions and the results of programming practice. Delete the MNIST dataset in codes before submit.
- Do not paste a lot of codes in your report (only some essential lines should be included).
- Pay attention to the efficiency of your implementation. Try to finish this homework without the use of **for-loops**, using matrix multiplication instead.
- Any open source neural network toolkits, such as TensorFlow, Caffe, PyTorch, are **NOT** permitted in finishing homework 1 except for data preprocessing.
- **Plagiarism is not permitted**.

1 Objective Questions (score 2)

These questions are all multiple choice questions. One or more options may be correct.

1.1 (score 0.5)

If you find your deep learning model has too large training error, what should you do?

A. Increase the number of parameters

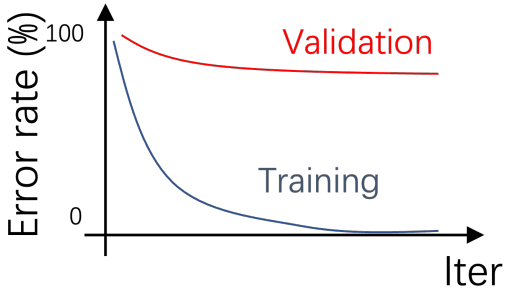
B. Increase the number of layers

C. Decrease the number of parameters

D. Add more regularization to the model

1.2 (score 0.5)

If you observe these curves during training a model, what should you try?



A. Increase the number of parameters

B. Decrease the number of parameters

C. Increase the regularization

1.3 (score 0.5)

Which claim about stochastic gradient descent (SGD) is incorrect?

A. It is beneficial for escaping from local minima

B. It's introduced to handle large training data

C. It slows down the training process

1.4 (score 0.5)

Consider the Softmax Regression model, which is often used in multi-class classification problems. Which of the following statements are true?

A. Softmax regression generalizes logistic regression to classification problems where the class label y can take on more than two possible values

B. The output of the softmax function is a vector that represents the probability distributions of a list of potential outcomes

C. The sum of the entries in the output vector from the softmax function is always greater than 1

D. Softmax regression can only be used for binary classification problems

2 Programming Practice (score 8)

2.1 Introduction

MNIST digits dataset is a widely used dataset for image classification in machine learning field. It contains 60,000 training examples and 10,000 testing examples. The digits have been size-normalized and centered in a fixed-size image. Each example is a 784×1 matrix, which is transformed from an original 28×28 grayscale image. Digits in MNIST range from 0 to 9. Some examples are shown below. **Note**: During training, information about testing examples should never be used in any form.

In this homework, you are required to use **Softmax Classifier** to perform MNIST classification.

Note: We will use Jupyter Notebook in homework 1, please install it.

2.2 Softmax for MNIST Classification

2.2.1 Files Description

There are several files included in the `./homework1-softmax/` folder:

- `homework1-softmax.ipynb` is an Jupyter Notebook file which describes the main contents of this homework. Data loading, hyerparameters setting, training and testing are included in this file. **Please read this file carefully**.
- `mnist_data_loader.py` is used to load MNIST dataset. Not required reading.
- `softmax_classifier.py` describes the softmax classifier. **You are required to complete** the function `softmax_classifier(W, input, label, _lambda)`. Part of this function is provided and you need to write down your code at **# TODO** to calculate the loss, gradient and prediction.

2.2.2 Requirements

You are required to complete the **# TODO** parts in above files. **You need to submit all codes and a short report** with the following requirements:

- Record the training and testing accuracy, plot the training loss curve and training accuracy curve in the report.
- The given hyperparameters maybe performed not very well. Modify them by your own, and observe how do they affect the classification performance. Write down the results and your observation in the report.