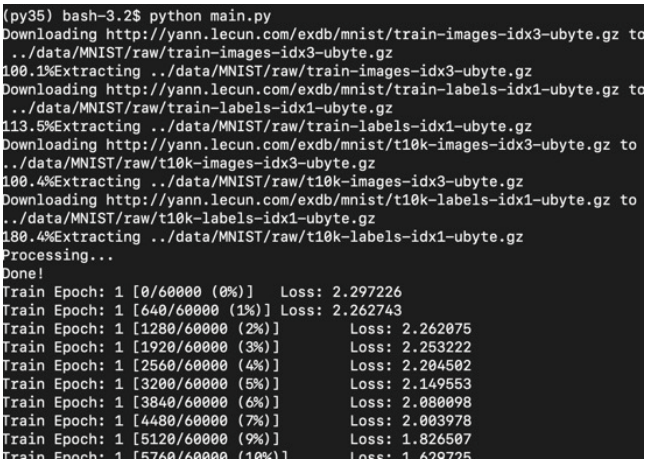
Spring 2019 Energy-efficient Machine Learning System

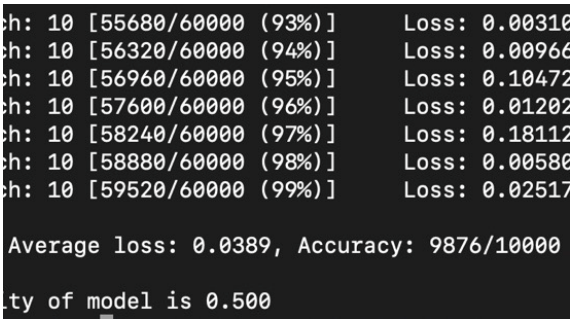
Homework Assignment 5

Due date: May 6th 2019

Problem (Prune a LeNet-5). In this problem, you are asked to train and test a neural network for LeNet-5 on MNIST dataset. Some information is as follows:

- A main.py file is already given, it is used to train a dense model, and test your pruned model
- An incomplete sparse_to_fill.py file is given. You need to fill some part of this file to implement pruning function
- When your filled sparse_to_fill.py file is correct, you only need to run main.py file, it should automatically train a dense LeNet, prune it and report the sparsity and accuracy similar to follows:





Performance Requirement and Submission:

- The test accuracy should achieve above 95%
- Submission should include your source codes and screen snapshot of your test accuracy after pruning and your sparsity ratio

Suggestion for hyperparameter setting (not necessary to follow): Check the default setting in the main.py file. You are allowed to change them

Hint: You can use some reference code in the slides of Lecture 12 if useful.