CS161: FUNDAMENTALS OF ARTIFICIAL INTELLIGENCE

Fall 2018
Assignment 9. Due Thursday, December 6, 2018, 11:55pm

Please submit your solutions on CCLE. The submitted file should be a **formatted PDF file** (no scans or pictures) containing **short answers to the questions** (you do not need to submit any code).

In this homework, we use a state of the art deep learning library to train a neural network model that classifies handwritten digit images. We provide a source file implementing steps required for solving the classification task: (1) loading the dataset, (2) visualizing training samples, (3) defining a network and an optimizer, (4) training the network, and (5) performance evaluation.

In order to complete this homework, you need to use Python 3, Jupyter, PyTorch, numpy, and matplotlib libraries. If you do not have enough familiarity with Python, we provide a virtual machine with these libraries installed as well as instructions on how to run the source file using the virtual machine. In this case, you need to install Virtual Box https://www.virtualbox.org/ and use a virtual machine image available on CCLE. You can find the soruce code on the virtual machine's desktop.

Please open the provided source file using Jupyter and read the notebook. Pay attention to the explanations at each code block and try to match them with the corresponding implementation. In order to read more about each function being used in the implementation, you may refer to PyTorch documentation at https://pytorch.org/docs/stable/index.html.

Plase use the provided implementation to answer the following questions:

1. What is the network architecture we used here? In other words, report the number of neurons at each layer, including input and output layers.

2. What is the activation function used in this network (please provide the name and mathematical function)?

3. What is the test accuracy achieved using this architecture?

4. Reduce the number of hidden neurons in each hidden layer by the factor of 16 ($size_{new} = \frac{size_{old}}{16}$), retrain the network, and report the test accuracy.

5. Compare the test accuracy of part 3 and part 4. Is there any change in the new test accuracy, why?

6. There is a variable named "MAX_ITERS", controlling the maximum number of training iterations. Use this variable and retrain the original network before the modification of part 4 for a 10 different MAX_ITERS values between 10 to 1000. Report the test accuracy for each. Do you see any trend in the achieved accuracy vs. number of iterations, explain the reason?