CS432 GPU Accelerated Computing
Assignment 2

CS Program
Habib University

Spring 2024
Due Date: 6 March 2024 @ 11:59PM

1 Introduction

This assignment has only one question. You are required to implement a multi-layered neural network [1] in CUDA. CPU implementation on Google colab in C++ using Eigen has been shared here: [2]. You should use this as reference when you do you CUDA implementation. All source files are available in the Assignment2 folder on the Github repo of our course [3].

2 Required components and rubric scores

Provide the following:

1. Implement a neural network class GPUNetwork in CUDA. (+20 marks)

2. Provide the following classes: GPULayer, GPUDenseLayer, GPUActivationLayer (+20 marks).

3. Provide the following activation functions on GPU: sigmoid, tanh and ReLU (+5 marks).

4. Provide the following loss functions on GPU: MSE, binary cross entropy loss functions. (+5 marks)

5. Implementation of XOR solver. (+20 marks)

6. Implementation of hand written digits detection using the MNIST dataset. (+20 marks)

7. Well written report containing training and validation loss graphs for XOR solver and hand written digits detection using MNIST using your GPUNetwork. Also elaborate the details about your CUDA based neural network design decisions. (+10)

3 Deliverables

Submit a colab notebook (.ipynb) file with your regn. no. as the filename i.e. $[ABxxxx\_Assgn1]$.ipynb. It should contain your solutions for the problems and the required work like graphs presented in a report form. Use the jupyter notebook text and code cells and write a consolidated report. **No other file name or file type will be accepted.** You might be called for a viva in case there is a need.

4 Deductions Note

Failing to comply to the instructions will result in a deduction of 20 percent score. Late submission policy as per syllabus will be applicable.

5 Using chatGPT or other AI software

You are not allowed to use any AI software to obtain the code for this assignment. **If you do use it, then you must share the chatgpt chat link**.

6 Plagiarism Policy

We have zero tolerance for plagiarism. The assignment submission should be your own genuine work without copying content from anyone else in the class or from the internet. If there is any evidence of plagiarsim, the case will be reported.

References

[1] Towards data science article, available online: `https://towardsdatascience.com/math-neural-network-from-scratch-in-python-d6da9f29ce65`. (Accessed on 10 Feb 2024).

[2] Student C++ MNIST reference Google colab notebook: `https://github.com/mmmovania/CUDA_Spring_2024/tree/main/Assignment2/Student_Reference_C++_MNIST.ipynb`

[3] Assignment 2 folder on the Github repo: `https://github.com/mmmovania/CUDA_Spring_2024/tree/main/Assignment2/`