CDA5125 Programming Assignment No. 3: OpenMP Deep Neural Network Code

(Due: 03/07/2022)

Purpose:

- Practice shared memory programming with OpenMP
- Practice parallelizing an application with OpenMP
- Experience tuning the performance of an OpenMP program on a NUMA node

Statement of work:

This is a group assignment. Each group can have 2 people. In this assignment, you will improve the deep neural network code that you developed in Assignment 2 by using OpenMP to explore parallel execution of the Deep neural network code. You would also tune the performance of your OpenMP deep neural network code in a NUMA node (linprog) by exploring the data placement and thread binding configurations.

Due dates:

The assignment is due on March 07, 11:59pm. Put all related source code, the makefile, and a README file in a tar file and submit the tar file. In the README file, you must describe (1) how to compile and run the program, (2) whether and how the parallelized program achieves high accuracy (parallelized code should still be correct), (3) report the speedups with vanilla OpenMP code, (4) describe the NUMA tuning that has been attempted and report the speedups after NUMA tuning. What is described in the README file must be repeatable with your submitted files.

Grading:

1. Submission has all components (all related source code, makefile, README file); the executable can be successfully produced with a 'make' command in the directory; a deep neural network for handwriting digit recognition with the MNIST dataset is built; OpenMP is used to parallelize the program (30 points).
2. The README file describes the information as required (5 points).
3. Points in 3), 4), and 5) can be obtained only after all points in 1) and 2) are obtained. One can follow the description in the README file to repeat the claims (5 points).
4. The improvement of the OpenMP program execution time is substantial (5 points).
5. The code runs significantly faster after NUMA tuning or your NUMA tuning is substantial (5 points).