COMP3046 Course Project – Part IV

Enhancement of your Artificial Neural Network (ANN)

Project Requirement: In this last stage of the course project, you are required to enhance your ANN by improving the speed of your training and reference functions, through one of the following techniques:

- OpenMP for multicore CPUs. You can also use AVX instructions and cache blocking techniques to further improve the performance.
- CUDA for Nvidia GPUs.

The key challenge in the implementation is to handle the whole mini-batch of data together so that you can transfer multiple matrix-vector multiplications into a single matrix-matrix multiplication. You may need to redesign your data structure. Please contact our TA Mr. WANG Qiang (qiangwang@comp.hkbu.edu.hk) if you have any difficulties.

Submission (through Moodle):

- A brief document that describes (1) the design of your solution; how do you optimize the performance; how to compile your source codes; (2) For Option 1: experimental results, and speedup analysis over the baseline version in Part III; You need to describe the hardware environment of your experiments; (3) For Option 2: a sequence of screen captures of your visualization; (4) the contribution of each team member (from 0 to 100%), including the design, implementation, debugging, experiments, documentation, and any other workload
- Source codes in a zip file
- A video demo of your project. The demo should show how to compile your project, how to train the model, and how to test the performance. Please make your demo concise.
- Deadline: **23:59PM, 30 Dec 2019**

Marking scheme (individual assessment based on each member's contribution):

- Parallel algorithm design: 20%
- Program correctness: 20%
- Performance (the shorter running time per epoch, the better performance you achieve): 40%
- Programming style: 10%
- Document: 10%