# HW2: Gradient Boosting

Weijie Zhao

10/04/2022

# HW 2: Gradient Boosting

- Given a training data A, a testing data B, a target testing accuracy C
- Train a gradient boosting model and output predictions

- 10  test cases. Each case weights 1 pt.
- The compilation is considered failed if it does not finish in 5 minute.
- A test case is considered incorrect if it does not finish in 2 minutes.
- Correct GPU solutions will get 5 pts bonus.
- The summation of the execution time across 10 cases will be uses to rank correct solutions.

- Due: 10/21/2022 11:59 pm EDT

# Testing Environment

- ssh yourusername@granger.cs.rit.edu
- Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz
- 48 threads in total (2 sockets, 12 cores per socket, 2 threads per core)
- 251 GB memory
- GPU: Tesla P4

- Testing limit:
  - 8 threads  taskset -c
  - 1 GPU

# Output Format

- N lines
- Each line contains an integer
  - The predicted class for each instance

# What Do We Need to Do?

- We are required to complete two scripts
- compiler.sh
  - it is executed once before the actual testing starts
- run.sh
  - it should takes 4 arguments
  - the first argument is the training data file name
  - the second argument is the testing data file name
  - the third argument is the target testing accuracy
  - the fourth one is the file name that you should write your results into

# HW2

- bunzip2 mnist.bz2
- bunzip2 mnist.t.bz2

- bash run.sh <train> <test> <acc> <out>
  - bash run.sh mnist.t mnist.t 0.9 sample1.out
  - bash run.sh mnist mnist 0.9 sample2.out
  - bash run.sh mnist mnist.t 0.9 sample2.out

  …
  - bash run.sh mnist mnist.t 0.97 sample10.out
- We guarantee that testing data will not have more features than the training data.