HW6

All 5 tests passed:

```
Output from main_q1
------------------
[==========] Running 5 tests from 1 test suite.
[----------] Global test environment set-up.
[----------] 5 tests from gtestTrain
[ RUN      ] gtestTrain.small1
Rank 0/1 started on node hpcc-gpu-5-3 [myGPU = 0, nGPUs = 4]
Time for Sequential Training: 0.0256574 seconds
Time for Parallel Training: 0.00529092 seconds
[       OK ] gtestTrain.small1 (178 ms)
[ RUN      ] gtestTrain.small2
Rank 0/1 started on node hpcc-gpu-5-3 [myGPU = 0, nGPUs = 4]
Time for Sequential Training: 0.00903242 seconds
Time for Parallel Training: 0.0116145 seconds
[       OK ] gtestTrain.small2 (35 ms)
[ RUN      ] gtestTrain.small3
Rank 0/1 started on node hpcc-gpu-5-3 [myGPU = 0, nGPUs = 4]
Time for Sequential Training: 0.0706677 seconds
Time for Parallel Training: 0.0915959 seconds
[       OK ] gtestTrain.small3 (177 ms)
[ RUN      ] gtestTrain.medium
Rank 0/1 started on node hpcc-gpu-5-3 [myGPU = 0, nGPUs = 4]
Time for Sequential Training: 0.0508043 seconds
Time for Parallel Training: 0.0409252 seconds
[       OK ] gtestTrain.medium (135 ms)
[ RUN      ] gtestTrain.large
Rank 0/1 started on node hpcc-gpu-5-3 [myGPU = 0, nGPUs = 4]
Time for Sequential Training: 0.337187 seconds
Time for Parallel Training: 0.140228 seconds
[       OK ] gtestTrain.large (710 ms)
[----------] 5 tests from gtestTrain (1237 ms total)

[----------] Global test environment tear-down
[==========] 5 tests from 1 test suite ran. (1237 ms total)
[  PASSED  ] 5 tests.
[anjiang@hpcc-gpu-5-1 starter-code]$ █
```

I used the option 1 to implement the loss reduction:

Two options—don't mix

bug

Option 1: divide λ by the number of processes; then

1. Calculate partial gradient: $-\dfrac{1}{N} \underbrace{\sum}_{\text{partial sum for GPU i}} \nabla_p \text{error\_function}(y_i, \hat{y}_i(p)) \left(-\dfrac{\lambda}{n_{\text{nproc}}} p \right)$

2. Call `MPI_Allreduce` across n_{nproc} processes.

Option 2: add regularization term after the `Allreduce`.

1. Calculate partial gradient: $-\dfrac{1}{N} \underbrace{\sum}_{\text{partial sum for GPU i}} \nabla_p \text{error\_function}(y_i, \hat{y}_i(p))$

2. Call `MPI_Allreduce`

3. Add regularization term: $-= \lambda\, p$

Be consistent

Stanford University