HW2 Review

- 22/22 submissions
- 6/22 correct solutions
- Fastest solution:
  - Prajjwal Mehta 162.56s
- Runner-ups:
  - Lincoln Mercuro 172.71s
  - Jesse Burdick-Pless 177.47s
  - Hanna Koh 204.7s GPU solution
  - Solutions no slower than 345.42s will get 15 pts

xxxtargzlog 8 245.64 [0.01, 0.01, 0.02, 0.04, 0.04, 0.04, 0.07, 120.0, 120.0, 5.41] [8, 9]

- All grades will be finalized at the end of 10/24

Projects 50 pts

• Cross-platform compilation	2
• High-performance implementation on CPU	5
• High-performance implementation on GPU	5
• Illegal input handling	2
• Multi-language support	1 for each language
• Non-trivial optimization/techniques	1 for each optimization
• Tasks: classification, ranking, regression, retrieval, clustering	1 for each task
• Documentation	2
• Benchmarking with baselines	5
• Presentation	10
• Demo	20

Required

Example Projects

- Toolbox of linear classifiers with kernel method support
  - including SVM, linear regression, and logistic regression
- Gradient boosting
- Deep learning framework
- Approximate nearest neighbor search framework (KNN)

HW 3: Tensor Library

- The task of this homework is to implement a high-performance tensor library with a python interface.
- For each case, the code is considered correct if and only if it finishes in 2 minutes; and the numerical error of each printed value is within 1e-3.
- We include 3 sample test scripts here. During the testing, the scale of other 7 cases will be no larger than these cases.
- Correct GPU solutions will get 5 pts bonus.
- pybind11-2.10.0 is pre-installed in the testing machine.
- Testing limit: 8 threads 2 GPUs

HW 3: Tensor Library

- No 3rd party code is allowed.
- 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 used to rank correct solutions.

- Due: 10/28/2024 11:59 pm ET

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
  - 2 GPU

pybind11

```
#include <pybind11/pybind11.h>
namespace py = pybind11;
int add(int i, int j) {
    return i + j;
}
#include <pybind11/pybind11.h>

int add(int i, int j) {
    return i + j;
}

PYBIND11_MODULE(example, m) {
    m.doc() = "pybind11 example plugin"; // optional module docstring

    m.def("add", &add, "A function that adds two numbers");
}
```

```
$ python
>>> import example
>>> example.add(1, 2)
3
>>>
```

```
m.def("add", &add, "A function which adds two numbers",  py::arg("i"), py::arg("j"));
```

```
int add(int i = 1, int j = 2) {
    return i + j;
}

m.def("add", &add, "A function which adds two numbers",
      py::arg("i") = 1, py::arg("j") = 2);
```