HPC Project Proposal
Final Project EECE5640 High Performance Computing
By Syed Shahbaaz Ahmed, Balaji Sundareshan

Problem Statement

The problem statement has two parts:

- To analyze Machine Learning(ML) and Deep Learning(DL) workloads on CPU, GPU and multi-GPU
- To use single-precision(FP32) and mixed precision(FP16 and FP32) for ML and DL workloads to obtain similar accuracy with reduced memory usage and faster training.

Workloads:

The following are the workloads that will be used in the project:

1. Linear Regression
2. Logistic Regression
3. Decision Trees
4. Image classification using Convolutional Neural Networks (CNN)
   CNN Architecture Backbone: ResNet - 18

Datasets:

Two types of datasets will be used. One with tabular data (for ML workload) and other, image (matrix-type) data.

- Datasets for ML:
    1. UCI Higgs Dataset (classification)
    2. UCI YearPredictionMSD (regression)

- Datasets for DL:
    1. CIFAR 100 Dataset (classification)

Platform:

- All the proposed experiments would be on Discovery Cluster (K40, V100 GPUs)
- **Framework**: Pytorch which uses C++ and CUDA in the backend, for parallel implementation in CPU and GPU respectively.
- **Programming Language**: Python

Evaluation experiments:

1. Analyze run time and model accuracy for training and prediction across GPU and Multi-GPU for single precision and mixed precision inputs for image classification workload.
2. Analyze run time and model accuracy for training and prediction across CPU, GPU and Multi-GPU for the following workloads:
    - Linear Regression
    - Logistic Regression
    - Decision Trees
3. Number of experiments:

Workload	Number of experiments
Linear Regression	(CPU, GPU, Multi-GPU) * (training, prediction) * (single precision) = 3 * 2 * 1= 6
Logistic Regression	(CPU, GPU, Multi-GPU) * (training, prediction) * (single precision) = 3 * 2 * 1 = 6
Decision Trees	(CPU, GPU, Multi-GPU) * (training, prediction) * (single precision) = 3 * 2 * 1 = 6
Image Classification (CNNs)	(GPU, Multi-GPU) * (training, prediction) * (single precision, mixed precision) = 2 * 2 * 2 = 8

Total experiments = **26** (6 + 6 + 6 + 8)

Extra credit (+20 on quiz):

- Image reconstruction using CNNs
    1. Analyze run time and model accuracy for training and prediction across GPU and Multi-GPU for single precision and mixed precision inputs.
    2. Total number of experiments = 2 * 2 * 2 = 8

Grading Scheme:

A+	Analyzed all 4 workloads for CPU, GPU and Multi-GPU. Total experiments = 26 (6 + 6 + 6 + 8)
A-	Analyzed 3 workloads for CPU, GPU and Multi-GPU. Total experiments = 20 (6 + 6 + 8)
B+	Analyzed 2 workloads for CPU, GPU and Multi-GPU. Total experiments = 14 (6 + 8)

References:

1. https://archive.ics.uci.edu/ml/datasets/HIGGS
2. https://archive.ics.uci.edu/ml/datasets/YearPredictionMSD
3. https://www.cs.toronto.edu/~kriz/cifar.html
4. https://pytorch.org/blog/accelerating-training-on-nvidia-gpus-with-pytorch-automatic-mixed-precision/
5. https://developer.nvidia.com/blog/gradient-boosting-decision-trees-xgboost-cuda/