MLDS 422 - Intro to Python
Lab 6

Sungsoo Lim

November 1, 2023

NORTHWESTERN
UNIVERSITY

Today's Lab Materials

- Multithreading
- Multiprocessing
- RAPIDS
  - CUDA based Python library for data analytics

Terminologies

- **Process**: An execution environment of a computer program (e.g. a Python script). Multiple processes can be running the same program, but they can use different data and compute resources.
- **Thread**: A unit of execution in a process. Threads only execute serially, but a process can have multiple threads running concurrently, taking on different parts of the task.
- **Multithreading**: Having the same process run multiple threads concurrently, sharing the same CPU and memory. Multiple threads cannot execute code simultaneously, but when one thread is idly waiting, another thread can start executing code.
- **Multiprocessing**: Multiple processes are spawn from the same process, each having its own CPU and memory. Concurrent processes can execute code simultaneously.

Multithreading - Advantages and Use Cases

- Response time, speed
- Input/Output in Python - downloading data from the internet and writing data to files
- Debugging can be more difficult

Multiprocessing - Advantages and Use Cases

- If it can be broken down into subprocesses that can run independent of one another
- Less overhead and more efficient debugging
- CPU bound tasks - multiple cores can be used simultaneously

Google Colab

- If you have a Google account, you can access GPU for free
- Runtime → Change runtime type → GPU
- https://colab.research.google.com

RAPIDS

- Use NVIDIA GPUs for data science Python commands
- https://rapids.ai
- Example - https://colab.research.google.com/drive/1jolXJ1BZOlWoOB-piOVnEqqLKLAslF7r