Deep Learning in Medicine

BMSC-GA 4493, BMIN-GA 3007, Spring 2020

Lab 4: HPC

B V Nithish Addepalli, bva212@nyu.edu

Today, we will go through the useful resources and requirements to set up your HPC account and environment which you'll need for your assignments as well as your course project.

- Refer to the link below for instructions for Tunnel Setup, to request compute and GPU resources, to transfer files between the HPC and your Local PC/Laptop:
    - A quick reference to access NYU High-Performance Computing
    - Logging into HPC
    - Official Wiki
    - For file transfers – easier and cleaner GUI (Globus)
        - Prince endpoint *(nyu#hpc)*
- Different storage options and the limits
    - Home – 20GB only!!
        - Don't download your project datasets here
    - Use locations like */scratch/<net_id>/* or */beegfs/<net_id>/* for data storage
        - Note that the data in these drives are flushed if not used for 60 days
        - Use */archive/<first_letter_of_net_id>/<net_id>/* for such long-term storage
            - Can't use these on compute nodes
            - So, only backup the important results/outputs/models etc. that you might want for future use
- Brief about various resources available and their performance
    - **GPUs:** 1080, K80 (8 or 12 GB), P40 (22 GB), P100 (16 GB), V100 (16 GB)
    - In increasing order of performance and queue time
- Loading existing modules shared between users
    - E.g. *Module avail python* (To show all modules pre-installed with python in their name)
    - *Module load python3/intel/3.7.3*
    - *Module unload python3/intel/3.7.3*
    - *Module purge*
- Setup your own environment following the below instructions to create a virtual environment and install custom packages
    - Virtualenv
        - *Module load python3/intel/3.7.3*

        - *virtualenv --system-site-packages <env_name>*
        - *source <env_name>/bin/activate*
        - *which python*
        - *pip install* required packages
        - No need to load Python module every time; just activate the environment
    - Conda
        - *Module load anaconda3/5.3.1*
        - *conda create -p /home/<net_id>/<env_name>/*
        - Activate the *<environment>*
        - *which python*
        - *conda install* required packages
        - **NOTE:** Have to load anaconda module every time and activate the environment
- Running the Jupyter Notebook on Prince
- Getting data onto Prince
    - Through *scp* or Globus
    - *wget* directly in the prince
    - Use Python package like *urllib* and download the data
- Giving folder/data access to others co-working on the project – File permissions
- For any queries, reach out to one of the TAs (Linux/Mac - Nhung/Nithish and Windows – Ravi) on Piazza or the HPC people by mailing to hpc@nyu.edu (They're really good, quick and responsive)