

Some Machine Learning Packages on Bridges

Joel Welling

Pittsburgh Supercomputing Center

Why Bridges? Relevant Machine Features

- Large memory means large training sets on RAM disk
- GPUs
 - 16 K80s now
 - 32 dual-NVIDIA nodes with Phase 2
- Large parallel jobs under MPI, etc.
- Local copies of large datasets
 - e.g. large chunks of Common Crawl

XSEDE

Extreme Science and Engineering
Discovery Environment



Why involve us?

- You can do your own builds of software, but if we install it...
- Better performance
 - e.g. MKL instead of Atlas
- Shared effort
- Version control

Some Packages

Excluding Hadoop/Spark tools...

- Theano
- Caffe
 - Mpi-caffe
- TensorFlow
- R
- MatLab
- Scikit Learn
- Weka?

Example: a theano job

```
#!/bin/bash -x
```

```
#SBATCH -N1 -t 60 -p GPU
```

```
module load python cuda theano
```

```
source $THEANO_ENV/bin/activate
```

```
python ${HOME}/theano/logistic.py
```

- This just runs the theano logistic regression example
- Requests 60 seconds on 1 node in the GPU partition
- Load the theano module, and other modules on which it depends
- Activate the Theano virtualenv
- Run the test

What's this 'module load' stuff?

```
[welling@br005 ~]$ module show cuda
```

```
-----  
/home/welling/git/bridges_modules/cuda/7.5:
```

```
module-whatis      CUDA 7.5  
conflict          cuda  
prepend-path      PATH /opt/packages/cuda/7.5/bin  
prepend-path      LD_LIBRARY_PATH /opt/packages/cuda/7.5/lib64  
setenv            CUDA_ROOT /opt/packages/cuda/7.5
```

```
-----  
[welling@br005 ~]$ module show theano
```

```
-----  
/home/welling/git/bridges_modules/theano/0.8.0:
```

```
module-whatis      theano-0.8.0  
prereq            cuda  
setenv            THEANO_ENV /opt/packages/theano/theano_0.8.0/theanoEnv
```

Run and results

```
[welling@br005 ~]$ cd /pylon2/pscstaff/welling
[welling@br005 welling]$ mkdir testdir
[welling@br005 welling]$ cd testdir
[welling@br005 testdir]$ sbatch /home/welling/theano/run_logistic.slurm
Submitted batch job 81251
```

```
[welling@br005 theano]$ squeue -u welling
```

JOBID	PARTITION	NAME	USER	ST	TIME	NODES	NODELIST(REASON)
81355	GPU	run_logi	welling	R	0:10	1	gpu004

... *time passes* ...

```
[welling@br005 testdir]$ squeue -u welling
```

JOBID	PARTITION	NAME	USER	ST	TIME	NODES	NODELIST(REASON)
-------	-----------	------	------	----	------	-------	------------------

```
[welling@br005 testdir]$ ls
```

```
slurm-81356.out
```