Suggestions for common ML software installations

Michael Kagan, mkagan@cern.ch

David Rousseau rousseau@lal.in2p3.fr

IML

September 28, 2016

Requirements

- We want the minimal threshold for using non HEP specific machine learning tools➔easy distribution
- See next page the list of packages from ATLAS survey
- Three targets:
  - Lxplus and any CVMFS enabled cluster/machine
  - Laptops (offline)
  - Non CVMFS enabled cluster/machine
- CVMFS
  - Some of it already exists (see http://lcgsoft.web.cern.ch/lcgsoft/release/85swan3/) e.g. :

```
$ source /cvmfs/sft.cern.ch/lcg/views/LCG_85swan3/x86_64-slc6-gcc49-opt/setup.sh
$ python
>>> from sklearn import linear_model
>>> import matplotlib.pyplot as plt
>>> ...
```

  - … would be nice to have it more nicely wrapped up lsetup ml
  - Need global versioning for reproducibility (is LCG_85swan3 sufficient ?)
- For laptops (Mac and Linux)
  - Michael K's 6 lines recipe (root hooks provided by Netherlands eScience Center, found by Gilles Louppe) for Anaconda+ROOT installation on ATLAS wiki, can be made widely available (screen shot attached to the agenda)
  - Installation scripts ?
- Non CVMS enabled cluster/machine
  - Lower priority. Need to understand how important is this target.
  - Michael K's Anaconda recipe seems to work

ML tools

The list collected so far in ATLAS (ML tool, or package required by ML) (need to specify versioning and source)

- Root numpy
- Numpy
- Hdf5, H5py
- Scipy
- Matplotlib
- Pandas
- Scikit Learn
- Scikit images
- XGboost
- Theano
- TensorFlow (note, this may require the google Bazel compiler)
- Keras
- BLAS / ATLAS / LAPACK libraries
- Cuda and CuDNN
- Zlib

- Not necessarily one mechanism to distribute them all