Data Mining using Python
— exercises for installation

Finn Årup Nielsen

DTU Compute
Technical University of Denmark

September 1, 2014

Install Python

Install python and some libraries and check that you can write:

```
$ python
>>> import numpy
>>> import scipy
>>> import cherrypy
>>> import sqlite3
>>> import nltk
>>> import sklearn
>>> nltk.download()
>>> from nltk.corpus import brown
>>> brown.words()
['The', 'Fulton', 'County', 'Grand', 'Jury', 'said', ...]
```

and that you can run a scikit-learn example (after downloading it)

```
$ python  plot_classifier_comparison.py
```

Install Python

Install `ipython` (e.g., by `pip`)

Start with:

```
ipython -pylab
```

Once installed make sure you can write:

```
In [1]: plot(sin(linspace(0,8,100)))
```

Install IPython Notebook

Check that you can run a IPython Notebook file.

Test that the IPython Notebook works with your own program or use, e.g., the "Modeling of Female Hurricanes dataset" example.

It should look something like this.

Web serving

After installing Cherrypy see that it works.

Try to get the `bonus-sqlobject.py` from the tutorial to work.

Note that this requires the installation of a SQL database. One of the line in the `bonus-sqlobject.py` file states:

```
# configure your database connection here
__connection__ = 'mysql://root:@localhost/test'
```

If you don't want to install MySQL try installing the simpler sqlite and its python support and then change the connection line to something like `'sqlite:///home/user/the_data_base_file.db'`

Installation tasks

Install spyder

Get a ‘Hello, World’ program up and running on a cloud service up, either with Google App Engine, Heroku or Pythonanywhere.

Git

Install git on you local computer.

Setup a git account in the cloud, e.g., at Github, Bitbucket or Gitlab account.

Note that you should be able to let us (lecturer, Teaching Assistants) have access to the repository. and that, e.g., Github requires payment for private repositories but may have educational discount.

Virtualenv

Setup two virtual environment with `virtualenv` and check that you can copy requirements between the two environments.