CM146, Fall 2017
Problem Set 0: Programming Environment Setup
Due Oct 09, 2017 at 11:59 pm

This semester, we will be using Python and scikit-learn, a machine learning package for Python. You should install the following software:

- python 2.7.x (https://www.python.org/downloads/)
- numpy (http://www.numpy.org/)
- scipy (http://www.scipy.org/)
- matplotlib (http://matplotlib.org/)
- scikit-learn 0.18.1 (http://scikit-learn.org/stable/)

The instructions for installing `scikit-learn` already include the instructions for installing `numpy` and `scipy`, so we recommend that you start there.

Alternatively, you can use a third-party distribution. For example, you can sign up for an Academic license for Enthought Canopy (https://www.enthought.com/products/canopy/) or Anaconda (https://store.continuum.io/cshop/anaconda/), which include these packages (though you may have to install them through the Package Manager).

There are a number of IDEs that make it easy to program in Python. We recommend PyCharm (https://www.jetbrains.com/pycharm/download/#section=mac) or Jupyter (comes pre-installed with Anaconda).

To test your setup, launch the Python interpreter from the command line. Make certain that it says that you are running version 2.7.x; if not, you may need to change the Python executable you are running.

To test `matplotlib`, run the following code in the Python interpreter:

```
import numpy as np
import matplotlib.pyplot as plt

x = np.arange(0, 5, 0.1);
y = np.sin(x)
plt.plot(x, y)
```

which should display a Matlab figure.

To test `scikit-learn`, run the following code in the Python interpreter:

```
from sklearn import tree
X = [[0, 0], [2, 2]]
y = [0.5, 2.5]
clf = tree.DecisionTreeRegressor()
clf = clf.fit(X, y)
clf.predict([[1, 1]])
```

which should give you the following output:

```
array([0.5])
```

To master python in just a few hours, the Google Developer Python tutorial is a good starting point (https://developers.google.com/edu/python/).

If you are familiar with Matlab, `numpy` and `matplotlib` should be very straightforward, and you can check out the handy guide on "NumPy for Matlab Users" (http://wiki.scipy.org/NumPy_for_Matlab_Users). If nothing else, you should look at the key differences.

If you want to play around with `scikit-learn`, you can also check out the official tutorial (http://scikit-learn.org/stable/tutorial/).