Installation

Python is one of the most popular language today for data science and Machine Learning. Apart from core python, we will be using additional packages which need to be installed separately.

As you work on different problems, you may need custom python environments with specific list of libraries. There are multiple ways to manage python environments. Popular tools/approaches are

- virtualenv
- dockers
- conda

Conda/Anaconda

We will use conda. Infact we will go one step further by insalling Anaconda which bundles python and hundreds of data science python packages together. For first part of the course we will do a default install.

Installation

Please visit link to download and install Anaconda. Please download Python 3.6 version, 64-bit version Graphical Installer for your operating system. The download size is about 400-500MB.

- link for Windows
- link for Mac
- link for Ubuntu/Linux

Pick all default settings and complete the install.

Anaconda comes bundled with a GUI that allows you to launch applications and easily manage conda packages,environments. However in this class we will stick to mostly using command line usage of conda for creation of envionrment and package management.

Testing your installation

1. After you have installed Anaconda, launch terminal

   - Windows 10 - type "cmd" in serachbox at bottom left and press enter.
   - Mac - Press CMD + Space to open spotlight search, and type terminal and hit return
   - Ubuntu - using key combination Ctrl+Alt+T

2. Type python at command prompt and hit enter. This will change the prompt to ">>>". You are now inside python shell.

3. Type "2+2" and hit enter. It will display the result "4"

4. Next type import numpy as np and hit enter. It will import numpy package and will bring back python prompt (>>>). Type np.random.random() and hit enter. It should display a number between 0.0 and 1.0

5. To exit the python shell, type exit().

Congratulations, you have python installed on your system alongwith a large set of data science libraries.

Our default installation comes bundled with Jupyter notebook which is a web application that allows you to create and share documents containing live code , visualizations and narrative text. This is the most preferred option when building you models. Code from jupyter notebook is converted to regular python script files before production deployment. We will be exploring various ways to writing python code in Session 1.