Numpy

Numpy is a numerical processing and computing package for Python, and one of the cornerstones of Data Science in Python. We will be using Numpy extensively throughout the course, even when we are using other libraries, which are almost definitely dependent on Numpy. To learn more about numpy:

1. Read <https://en.wikipedia.org/wiki/NumPy>
2. Read <https://docs.scipy.org/doc/numpy-dev/user/quickstart.html> and try it out as much as you can. It’s okay if you can’t finish it, because we will cover everything in our course.

Machine learning, unsupervised learning, clustering and K-means clustering

Machine learning is arguably the most important component of data science. Whether you are new to these concepts or not, read the following pages so that you have a basic understanding of what we are trying out in the first class:

1. <https://en.wikipedia.org/wiki/Machine_learning>
2. <https://en.wikipedia.org/wiki/Unsupervised_learning>
3. <https://en.wikipedia.org/wiki/Cluster_analysis>
4. <https://en.wikipedia.org/wiki/K-means_clustering>

Jupyter notebook

Jupyter notebook (formerly IPython notebook) is an interactive web interface to write and execute Python code that we are using throughout the course. To learn more:

<http://bebi103.caltech.edu/2015/tutorials/t0b_intro_to_jupyter_notebooks.html>