P556: Applied Machine Learning

Programming Environment

The language of choice for all the assignments in this class will be Python 3. Moreover, we will be working with the Anaconda distribution which not only contains the data science and machine learning Python packages but also a variety of tools and editors which are very helpful. You can install the distribution for your system from this page. As you must know, Python has a number of libraries or packages which are used for different tasks across data science. For instance, pandas is used to work with datasets whereas scikit-learn contains implementations of a large number of machine learning algorithms. Many times, certain packages require specific version of certain other packages and managing the installation of all these packages can get very tedious. The `conda` package manager included with anaconda distribution can make your life a lot easier. So in short Anaconda comes is a bundle of the following things:

1. Useful Python packages (numpy, scipy, pandas, etc.)
2. Package Manager to manage the above packages (conda)
3. Tools (Jupyter qtConsole, Jupyter Notebooks, Spyder, etc.)
4. Anaconda Navigator that is GUI to manage all the above things (Only available for Mac and Windows. You have to use command-line on Linux, which is way more cool).

In this document we will provide a set of resources that will help you get familiarized with the all the necessary tools that are required to become adept in using Python. We realize that all this looks daunting in the beginning, but with practise, it will definitely increase your productivity. Whenever encourage your to read as much of the official documentation as well. Do not limit yourselves to the tutorials which we provide. This is not meant to be a comprehensive resources for everything that will be covered in the course, but rather a starting point to help you not get overwhelmed. We might miss certain things. 'Google' will be your best friend.

In order to not get overwhelmed, we recommend learning things in the following order. You are obviously encouraged to deviate and explore on your own as well.

1. Install Anaconda
2. Learn about the conda package manager.
3. Learn what Virtual Environments are and create one for this class called 'aml'. Make sure that you do all your assignments in this environment.
4. Familiarize yourself with 3 important tools:
    a. Jupyter qtconsole
    b. Jupyter Notebook
    c. Spyder - This is an IDE which comes with Anaconda
5. Learn the basics of the following Python packages:
    a. Python list comprehensions and Lambda functions

    b. Numpy
    c. Pandas
    d. Matplotlib
    e. Scikit-learn*
    f. Scipy*
6. Finally, you should learn how to write Math equations using Markdown. Markdown uses LaTeX syntax for writing math. This is how you will be writing any math equations in the assignments in the class (Not by hand!).

* You do not need these packages for Assignment 0, but will need them for the rest of the class.