Getting Started with Python and NumPy

Complete your work in a Jupyter Notebook and submit a PDF of the results to Canvas.

Machine learning software packages have advanced dramatically in the last ten years, and Python has become the default and primary language for machine learning. This first activity will go over a few Python basics.

The Jupyter Notebook is an open-source web application that allows you to execute scripts and view results and plots in a browser. We will rely on both Python and Jupyter notebooks extensively in this course.

1. Set up your environment (skip this if you've completed the first activity). There are several options for getting up and running:

   - Install Python and Jupyter on your laptop. This is the preferable approach, but takes more effort.
   - Use the Computer Aided Engineering (CAE) Jupyter setup (the link can be found on the course website).
   - Use a cloud/web environment such as Google Colabs.

2. Lists, tuples, sets and dictionaries. Use the starter notebook to help with the remaining problems. Four built in data structures in Python that we will make use of are lists, tuples, sets and dictionaries.

   a) A list is exactly that: a list. Python lists are ordered and changeable. Write code to convert the following paragraph to a list of strings:

      this deluge of data calls for automated methods of data analysis which is what machine learning provides in particular we define machine learning as a set of methods that can automatically detect patterns in data and then use the uncovered patterns to predict future data or to perform other kinds of decision making under uncertainty

   b) Create a new list that consists of every third word from the original list. Do not write a for loop.

   c) What is the 16th from last item in the list?

   d) Convert the list to a tuple. When should you use a tuple instead of a list?

   e) Convert the list to a set. How many items are in the set? What is the difference between a list and a set?

f) A Python dictionary consists of *key, value* pairs. Create a Python dictionary where each distinct word in the paragraph above is the key, and the corresponding value is the count of times that word appears.

g) Which word/words appears most frequently, and how many times does it appear?

h) Use a *list comprehension* to convert the dictionary to a list of tuples, where the first element of the tuple contains the key, and the second element contains the value. Note that your code should be a single line.

3. NumPy is a numerical computing package for Python. You'll often see the library imported at the beginning of a script as `import numpy as np`.

a) Python lists can be converted to NumPy arrays. Manually create a *list of lists* that represents the matrix

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 4 & 4 & 4 \\ 5 & 6 & 9 \end{bmatrix}$$

b) Convert the list of lists to a numpy array using `np.array()`.

c) What is A^{-1}? You may find `np.linalg.inv()` helpful.

d) Create a length n array random numbers distributed uniformly between 0 and 1 using `np.random.rand()` for $n = 10$.

e) Repeat for $n = 10, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8$. For each value of n, record the minimum value of the array. *Hint: you can do this in one line with a list comprehension.* Create a plot with n on the horizontal axis and the logarithm of the maximum value on the vertical axis. Are the results what you expect? Why?