Lab 01 – Getting started

This lab has two purposes: first making sure you are all setup and ready to work in the labs or your own computer; and second, getting you working on some basic examples of NLP.

Coding IDE Options

There are several options on how you can get support when coding:

1. Using a text editor and a command line terminal to run the python files. This is not recommended, since is counterproductive (and very frustrating).

2. Using a Python IDE such as Visual Studio Code (https://code.visualstudio.com) or PyCharm (https://www.jetbrains.com/pycharm) and many more. Those are great tools that are fully integrated and support most needed functions (including running Jupyter Notebook).

   The main drawback is that, although you have full control on the setup and installation, they depend on the capability of the machine that are installed. So, if your laptop/desktop is not powerful enough, you will be struggling to run some of the more demanding ML code (e.g. if GPU is needed).

3. Another option, and the **most recommended**, is to host and run the Python code in the Cloud while accessing and interacting with the code remotely via a Web interface. In other words, you will be using your standard browser to access a Web-based IDE that will be running on a machine in the Cloud. Most popular platform are:
   a. Colab by Google (https://research.google.com/colaboratory)
   b. recently AWS also launched Studio Lab (https://studiolab.sagemaker.aws)

   Both platforms are completely free and allows you access to hardware that is capable to run ML code via CPU or GPU. Those IDEs are based on the Jupyter Notebook and therefore you can run your code and document your experiments in the same interface. The only notable drawbacks of a Cloud IDE are that, first, that they require an Internet connection, and second, those IDEs are free because the hardware resources are limited by time or capability… although good enough for our purpose of learning.

4. The last option, a very common one actually, is to install the Jupyter Notebooks on your own computer, using Anaconda (https://www.anaconda.com).

Development Environment Setup

1. If you are going to use your own machine or the labs, make sure Conda is installed and working (should be in the labs)
    a. Try on a terminal or command window (no need to do that for the cloud IDE):
    ```
    % conda --version
    ```

    b. If Conda is not installed (i.e. might not be on your laptop), then follow the guide to install it:
        i. instructions here (using Anaconda, recommended): https://docs.anaconda.com/anaconda/install
        ii. or here (bit more advanced, so try Anaconda first): https://docs.conda.io/projects/conda/en/latest/user-guide/install/

2. Setup a Python environment to work in (e.g. `nlp2022`) using a command such as:
    ```
    $ conda create --name nlp2022 python=3.9
    ```

    Setting up an environment is important for ensuring you have the same installation as everyone else you work with and for preventing some versioning issues. This will also help with your coursework when you work in a group.

    Follow instructions here:
    https://docs.conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html

    If you are setting up the environment on Studio Lab, then you will need to run the necessary commands in the provided terminal window. See instructions here:
    https://docs.aws.amazon.com/sagemaker/latest/dg/studio-lab-use-manage.html

3. Enter this environment before running the labs!
    ```
    $ conda activate nlp2022
    ```

Your first NLP lab

1. Access the first lab on GitHub at:
   https://github.com/surrey-nlp/NLP-2022

2. Clone the repository into a location accessible by the IDE you chose to use.
   `git clone https://github.com/surrey-nlp/NLP-2022.git`

   For help check: https://docs.github.com/en/repositories/creating-and-managing-repositories/cloning-a-repository

   This repository will be updated each week with more examples.

   Alternatively, you can also download the entire code repository as a zip file, but is not recommended for the long term.

3. If you are running the Jupyter Notebook in your local machine or in the labs, make sure you have created and loaded the right Python environment first. If you are using a Studio Lab, then start a new terminal window and create the environment there.

4. Start a Jupyter Notebook and open `lab01.ipynb`
   a. To run the notebook, see instructions here:
      https://jupyter.readthedocs.io/en/latest/running.html

   b. You could also try JupyterLab for a desktop like IDE like experience (if you are not using the Cloud IDE):
      https://jupyterlab.readthedocs.io/en/stable/getting_started/starting.html

5. Now, just follow the instructions in the notebook