

W4705 - Natural Language Processing
Homework 0 (Warmup)
Introduction to the Programming Environment
Due: 12th September, 2017

Introduction

This assignment has 2 goals:

1. Give you an idea of the background needed for the class, and
2. Help you get familiar with the environment in which the assignments will be created and submitted.

Environment Setup

Step 1: Spawn an instance following the instructions mentioned in [“Preparing Your Google Cloud VM for W4705”](#) document

Step 2: SSH into your instance

Step 3: If you are unfamiliar with a Linux/Unix based system, go through the basic commands [here](#)

Step 4: Install numpy: `sudo apt-get install python-numpy`

Step 5: Install scikit-learn: `sudo apt-get install python-sklearn`

Assignment

HW0_starter.py file, accompanying this document, includes code for scikit-learn

KNeighborsClassifier on 20newsgroups data. Run the code and observe the output (Ignore “No handlers could be found for logger” message, if received).

Your assignment is to plot the trend of average F1-score (i.e. number-of-neighbors vs F1-score), as the number of neighbors in the KNeighbors Classifier increases from 2 to 50. Use `sklearn.metrics.f1_score` with `average = 'weighted'` for getting the average F1-score, and `matplotlib` to plot the graph.

Submit your assignment

The following files are required to be turned in for this assignment:

1. Python code which outputs the plot of number-of-neighbors vs corresponding F1-score.

Submit your assignment to the BitBucket repository following instructions mentioned in the [“How to submit assignments”](#) document.

Grading criteria

The assignment is worth 2 points. 1 point if the Environment setup is correct and 1 point for the python code which outputs the plot correctly.