DS2001 - CS Practicum

Spring 2023

Practicum 6 – Dictionaries

Practicum is DUE at the end of your scheduled practicum.

We'll be working in pairs, and today you can choose your own partner instead of being assigned. You may not finish every problem every week, but we expect your best effort.

We will let you know when there are 15 minutes left in class – at that point, it is time to wrap up your code and move on to the self-reflection portion of the assignment! Read the self-reflection questions below, write your answers, and save them in a PDF.

Submit both your code and your self-reflection PDF on Gradescope.

Grading Policy
You will receive full credit if you submit your self-reflection (with thoughtful answers) and any code by the deadline. All questions in the self-reflection must be answered for full credit.

Feedback
We will provide feedback on your code if requested in your self-reflection. Otherwise, we will not look at your code. Our feedback will mirror the expectations of your DS200 homeworks, and we will use the same criteria for grading your projects in DS2001. So, make sure you ask for feedback when helpful, and read our notes!

Today's Goals

- Create word count dictionaries for song lyrics
- Predict the writer of a sample lyric using word count dictionaries

Working in Pairs

Each pair will work at one computer, and everyone contributes. Here's how we split up the work:
- Navigator: Dictates the code to be written. Explains the *why* as we go. Checks for syntax errors.
- Driver: Writes the code. Listens closely to the navigator. Asks questions whenever there is a lack of clarity.
- Both driver and navigator are responsible for contributing to the work, and for making space for the other person to make contributions.

Programming Assignment

1. **Gather Data**

   Download beyonce.txt and taylor.txt from the course website. These files contain song lyrics for Beyonce and Taylor Swift, respectively. We will use these files and dictionaries to create a program that can predict which artist is more likely to have written a given lyric.

   To start, write a function to read in a file and return a list of strings in the file. Each element in the list should be a single word. The files you're using today contain multiple words per line, but do not contain any punctuation or special characters.

   Ex:
   > If the function is called with this file:
   > ```
   > data science is great
   > this practicum is great too
   > ```
   >
   > It should return the list:
   > ["data","science","is","great","this","practicum","is","great","too"]

   Call your function in main to create a list for Beyonce, and to create a list for Taylor.

2. **Computations**

   ## Word count dictionaries

   Write a function to turn a list of words into a word count dictionary. In the dictionary, keys should be words and values should be the frequency of each word in the list.

   Ex:
   > If the function is called with the list:
   > ["data","science","is","great","this","practicum","is","great","too"]
   >
   > It should return the dictionary:
   > {"data":1,"science":1,"is":2,"great":2,"this":1,"practicum":1,"too":1}

   ## Prediction

   Write a function with 5 parameters: (1) Artist 1's name, (2) Artist 1's word count dictionary, (3) Artist 2's name, (4) Artist 2's word count dictionary, (5) A song lyric. The function should split the song lyric into individual words, count how many times each individual word is used by artist 1 and by artist 2, sum the counts for individual word usage for each artist, and return a prediction that the artist with the higher sum wrote the song lyric.

3. **Communication**

   In `main`, prompt the user for a lyric. Use the functions you wrote above to tell the user who probably wrote the input lyric.

   Ex:
   ```
   Enter a lyric!
   calm down and say my name
   That lyric was probably written by: Beyonce
   ```

At this point, please complete the self-reflection! If you still have time afterwards, scroll down for a little extra/bonus work you can do on this assignment.

Self-Reflection

Create a PDF with answers to the following questions (each should be answered collaboratively with your partner unless instructed otherwise):
- Does this work reflect the best effort of both partners?
- In what way(s) did each partner (1) make contributions to the assignment, and (2) give the other person space to make contributions? (Each partner should respond to this question separately.)
- Which problem(s) would you like feedback on and why?

Keep Going (if time!)

Here are some ways you can improve and build on what we've done so far. Try any/all of them in the time we have left.
- The program we described is pretty specific to having exactly two songwriters. Try making it more generic and flexible by having your predictor function accept a list of wordcount dictionaries, instead of two individual ones, as well as a list of artist names.
- Did you write Part 3 such that the user could enter uppercase and/or punctuation? If not, add it now!
- Pick a third songwriter you like and create a text file for them (we created the Taylor and Beyonce ones by simply googling their lyrics), clean it up to remove punctuation and turn everything lowercase, and add them to your predictor model.