

Homework 2 - Linear Algebra and Probability Review

Dorian Goldman

January 6, 2017

1 PROBLEM 1 - SET UP GITHUB

- Go to <http://www.github.com> and create an account.
- Send your Github username to dg2991@columbia.edu.
- Install Git - <https://github.com/blog/1510-installing-git-from-github-for-mac>.
Make sure to install command line tools.
- When I have received your email, you should get a confirmation that you have been added to the repo.
- Click on this link:
<https://classroom.github.com/assignment-invitations/11415026d0459793405d3c1ff95cc259>
- Follow the instructions to clone that repo to your local machine.
- You should type a command like: `"$ git clone https://github.com/Columbia-Intro-Data-Science/python-introduction-(your-github-username).git"`

2 PROBLEM 2 - WRITE A PYTHON SCRIPT AND PUSH IT TO THE GITHUB CLASSROOM REPO

- If you haven't already, install Anaconda on your machine, which gives you iPython notebook: <https://www.continuum.io/downloads>

- Create a file called test.py in the folder where you've created your new repo (for the assignment, not the course).
- Make your script test.py return the product of two numbers, entered via the command line. **Eg: \$ python test.py 3 4** should return **12**.
- Add this file to the repo python-introduction-{your-github-username} by typing (make sure you are in the directory python-introduction-{your-github-username}) :
- \$ git add test.py
- \$ git commit -m 'Added my homework'
- \$ git push origin master.

3 PROBLEM 3 - CREATE A NOTEBOOK THAT READS IN A CSV FILE WITH PANDAS

In this exercise, you will create an iPython notebook. You will need to have installed Anaconda. To begin, open up Anaconda and launch Jupyter notebook.

- Within the Jupyter notebook, import the pandas library and load in the csv here:
`https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data`
- Using the groupby clause within pandas, find the percentage of people who are making above \$50k according to their *gender, race and country*.
- Push your notebook to the repo python-introduction-{your-github-username} following the steps you took in Problem 2.

If you need help: Here is a sample notebook you can use to get started: https://github.com/doriang102/Columbia_Data_Science/blob/master/homework/Homework-1-help.ipynb