VirginiaTech

Text mining Homework

For each assignment, turn in by the due date/time. Late assignments must be arranged prior to submission. In every case, assignments are to be typed neatly using proper English in Markdown.

This week, we spoke about Python. Python is an important language to know as a data scientist, in some disciplines, it is the dominant language, e.g. Deep Learning. In particular, Python 3. Our homework this week will be two problems as follows in Python 3.

Problem 1

You will implement a text mining algorithm in R and create as output a word cloud. Included in the folder 11_Python is a text file constitution.txt. Please implement a fully functional Rscript that loads, parses and finally creates a png image of the wordcloud. Hint: the R libraries tm and wordcloud may be useful. You may also want to look up "stopword" as a way to make the wordcloud more interesting. The script should be runnable as:

Example

```
$ Rscript wordCloud.R
```

Problem 2

Now, port your R code for creating a wordcloud to python. The python packages BeautifulSoup for text mining (may be overkill here) and wordcloud for the wordcloud.

Example

```
$ python wordCloud.py
```

Problem 3

In an earlier homework, you coded up a gradient descent algorithm to compute the roots of a function. Here, you should port that code over to Python. Please choose sensible bounds to get a single root and create a plot of the function along with the descent trajectory. You will may want to use matplotlib or ggplot2.

Example

```
$ python gradientDescent.py
```