

Assignment 1

L715/B659

Due Tuesday, September 6

1. The first part of your first assignment is going to be to browse the literature on a topic and find papers that you're interested in reading, using the syllabus as a starting point.

More specific notes:

- I encourage you to chat with me and with each other about topics, papers, etc. over email, before/after class, in my office, and so forth.
- Note that, although I'd like to get some spread over topics, multiple people can cover the same general issue, seeing as how there is a lot of depth to these topics, as well as some readings that are more foundational topics vs. ones that are more cutting-edge. Working together on leading a topic discussion is also allowed.
- You won't be exhaustively researching a topic (and you won't understand everything you skim right now), so think about what it is that you find interesting: particular subtasks? incorporating linguistic features? gathering data? combining different classifiers? using unsupervised learning? etc.

This assignment is not about committing to a particular research topic, but it doesn't hurt to be thinking about the types of projects you'll eventually be interested in working on.¹

- On the 6th, I would like you to turn in an approximately two-page report on the topic of your choice, what you hope to discover in that topic, a list of about 5–10 readings that seem to be most relevant to that hope, and any other notes you deem worth reporting.
 - On either the 6th or the 8th, we will then go about figuring out the exact order of discussions and scheduling meetings with me to talk about the topics.
2. The second part is to get you prepared for working with machine learning software. For now, we're all going to use Python (3) (<http://python.org>), so install the following:
 - (a) NumPy (<http://www.numpy.org>) and SciPy (<https://www.scipy.org>)
 - You could consider using a package manager like Anaconda (<https://docs.continuum.io/anaconda/>)
 - (b) pandas (<http://pandas.pydata.org>)
 - (c) BeautifulSoup (<https://www.crummy.com/software/BeautifulSoup/>)
 - (d) NLTK (<http://www.nltk.org/>)
 - (e) scikit-learn (<http://scikit-learn.org/>)

Again, talking with each other may help, in case you encounter difficulties.

¹More on projects in probably a month or so.