

Lab #8

Agenda

- Midterm
- Projects
- Additional built-in Python libraries

Projects

- Apply Python in a large scale project than class assignments
- **Proposal:**
 - Meant to just be a sketch of the project idea to make sure it is adequately coped
 - Due: Oct 31st (this Thursday)

Projects

1. Modeling Project

- a. Using a (e.g. machine learning) model, extract new insights from a data set

2. Visualization Project

- a. Use visualization methods to display/show interactively data sets in an informative way

3. Tooling Project

- a. Build / extend a Python library to perform some functionality relating to data science

My Protips

Don't do a project related to financial markets prediction.

Don't do a project related to financial markets prediction.

Don't do a project related to financial markets prediction.

Only do something related to deep learning if you feel comfortable with the basic concepts/libraries.

1. Modeling Projects

Example 1:

- Predicting credit card fraud
- Data set:
 - “284,807 credit transaction by European cardholders in September, 2013”
- Metrics:
 - Accuracy, recall, F1
- Model:
 - Logistic regression, random forests, XGBoost (gradient-tree boosting)
- Additional challenges:
 - Addressing the large imbalance in data distribution

1. Modeling Projects

Example 2:

- Predicting Flight Delays using geospatial and weather data
- Data set:
 - Grabbed data using APIs
- Metrics:
 - ?
- Model:
 - Random forests, XGBoost (gradient-tree boosting), etc
- Additional challenges:
 - Scraping the data

1. Modeling Projects

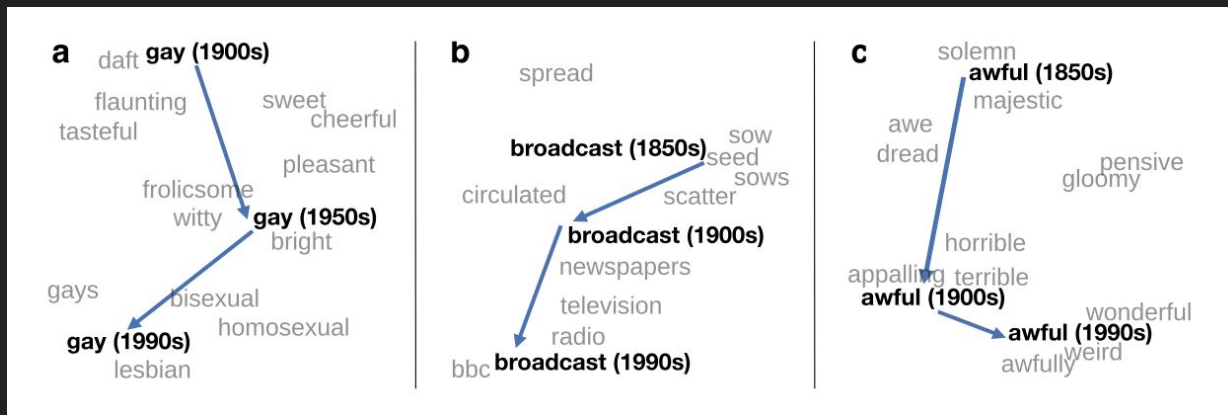
Example 3:

- Identifying Social Bias in Word Vectors
- Data set:
 - GloVe word vectors
- Setup:
 - Train or use an existing text sentiment classification model
 - Given a set of nationalities (e.g. “British”, “Ethiopian”), find the N closest words in the embedding space, and compute the average sentiment score of the neighboring words, and test for statistical significance in differences

2. Visualization Projects

Example 1:

- Visualize the evolution of word vectors over time

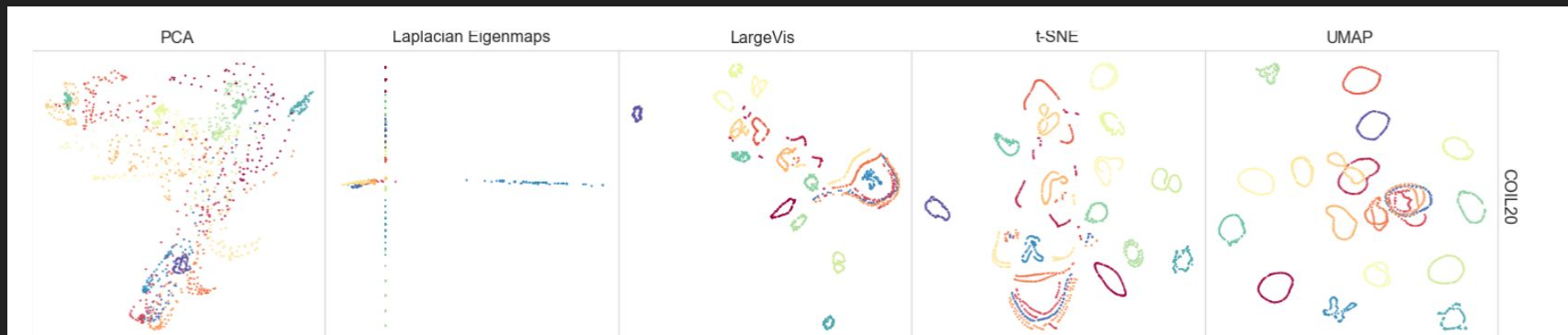


2. Visualization Projects

Example 2:



- Comparison of visualization methods for a data set and the apparent insights



3. Tooling Projects

Examples:

- Developing a web-scraper for a specific website and/or exposing it as an API
- Developing a library for working with/manipulating a specific kind of data (e.g. social media network graphs, database schema for e-sports data)
- Developing a web application for interactive visualization/exploration of a dataset