

Inference and Representation, Final Project

Proposal Due 10/24 by Email
Report Due 12/16

October 3, 2016

- The course final project is about applying tools you have learnt during the course to a specific dataset and scientific question. It is a research project that can have a scope outside the course, e.g. it could be the seed for a longer research project.

The first step is choosing a domain and dataset. These are some guidelines that may help:

- Not too small (at least 100 observations).
 - If you study a time-series, it should be approximately stationary, or made easily stationary by appropriate transformations.
 - Unique! Try to find a dataset/question that particularly interests you.
 - Have a look at the course syllabus, the project can be grounded on techniques that have not yet been covered in class but that will be in the upcoming lectures.
- **Proposal:** The proposal, due 10/24, must detail the question you are planning to address, which dataset are you planning to use and how it will be processed, which family of machine learning algorithms will be applied, and how are these methods going to be evaluated. Email the proposal to bruna@cims.nyu.edu and sontag@cs.nyu.edu. We will try to give constructive feedback on your proposal before spending time on the analysis.
 - **Note:** if you are interested in Natural Language Processing applications and are also taking “Natural Language Processing with Distributed Representations” by Prof. Sam Bowman, we are open to the idea of a larger, common final project. The proposal will need to be approved all parties.
 - **Some resources:** Here are some resources with plenty of available dataset. However, we encourage you to find on your own field of interest/expertise.
 - <http://cs.nyu.edu/~dsontag/courses/ml16/assignments/projects.html> contains a list of links to publicly available datasets.

- <https://wrds-web.wharton.upenn.edu/wrds/>: Financial data (requires access, talk to Joan B. if you are interested).
- Yahoo Finance
- World DataBank <http://databank.worldbank.org/data/home.aspx>: Socio-economic indicators.