

# Inference and Representation, Final Project

Proposal Due 11/13 by Email  
Report Due 12/12

- The course final project must be done in groups of 2 or 3 students.
- The final project can take any of the following forms:
  1. In-depth review of one topic, not covered in class, based on several recent papers related to some of the contents of the class. Think of this as a small research project.
  2. Open-Source, high-quality software implementation of one or several algorithms related to the class that currently is not publicly available,
  3. New application of a machine learning model to a particular dataset and to a specific scientific question,
  4. Research project that addresses a machine learning method or a theoretical question. The research project can have a scope outside the course, e.g. it could be the seed for a longer research project.
- If the project concerns an application, these are some guidelines that may help when it comes to choosing the dataset and the question:
  - 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 books. The project can be grounded on techniques that have not been covered in class.
- **Proposal:** Write a 1-2 page proposal that details 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 [soledad.villar@nyu.edu](mailto:soledad.villar@nyu.edu). I will try to give constructive feedback on your proposal before spending time on the analysis.

- **Presentation:** There will be a 10 to 15 minute presentation to be scheduled on December 4-6 afternoon/evening. Together with your proposal indicate your group's availability to present your work (date and time). Everyone is welcome to the presentations. The presentation should motivate the problem or goal, explain the method, its fundamental ideas and main applications. Questions will be asked and it's expected of students to be able to explain the concepts in depth.
- **Some ideas:**
  - Applications of Gaussians Process to data science.
  - Practical uses of HMM.
  - Comparison of generative models from GANs, normalizing flows, VAEs, etc.
  - Connection between belief propagation and message passing neural networks.
  - Latent Dirichlet Allocation.
  - Gibbs sampling in data science.
  - Exponential families and relevance for inference.
  - Inference techniques for domain adaptation.
  - Gerrymandering
    - \* Uses of MCMC to detect Gerrymandering  
<https://sites.duke.edu/quantifyinggerrymandering/>
    - \* Implementation of neural network to play Utility Ghost on real maps <https://arxiv.org/abs/1812.07377>.
  - Last year, Joan Bruna created a page with some resources and bibliography that you may use as inspiration for your project: <https://docs.google.com/document/d/1Ywj99qfrJ06g7CU3qi0Gar93GqaqhTK61LHiyJVwew/edit?usp=sharing>.
  - New ideas are welcome.