Introduction

In this course, you will learn advanced topics in machine learning. Part of the learning will be through in-class lectures and take-home assignments, but you will really gain hands-on experience with machine learning in your final project. We would like you to choose wisely a project that fits your interests. One that would be both motivating and technically challenging.

Most students do one of four kinds of projects:

- **Theoretical project.** Analyze and prove some interesting properties of a new or an existing learning algorithm.
- **Algorithmic project.** Develop a new learning (inference or optimization) algorithm, or a novel variant of an existing algorithm, to solve the learning problem.
- **Modeling project.** Develop a new neural network architecture, or a novel variant of an existing model, to solve the learning problem.
- **Application project.** Pick an application of interests, and explore how best to apply learning algorithms to solve it efficiently.

Many fantastic projects come from students picking either an application area that they're interested in, or picking some subfield of machine learning that they want to explore more. So, pick something that you are passionate about! Be brave rather than timid, and do feel free to propose ambitious things that you're excited about. (Just be sure to ask us for help if you're uncertain how to best get started.) Alternatively, if you're already working on a research project that deep learning might apply to, then you may already have a great project idea.

Hints

A very good project will be a publishable or nearly-publishable piece of work. Two of the main machine learning conferences are ICML and NeurIPS. You can find papers from recent ICML conferences online: http://icml.cc All NeurIPS papers are online, at http://books.nips.cc/. You can browse some of the recent machine learning papers to get inspired. You may also want to look at class projects from projects from Stanford CS230: https://cs230.stanford.edu/past-projects/

Evaluation

Projects will be evaluated based on:

- **Novelty of the work.** Is this project applying a common technique to a well-studied problem, or is the problem or method relatively unexplored?
- **Significance of the solution**. Did the authors choose an interesting or a "real" problem to work on, or only a small "toy" problem? Is this work likely to be useful and/or have an impact?
- **Technical quality of the work**. Does the technical material make sense? Are the things tried reasonable? Are the proposed algorithms or applications clever and interesting? Do the authors convey novel insight about the problem and/or algorithms?
- **Presentation and writing.** Is the idea and the solution clearly conveyed? Are the figures and tables carefully crafted? Is the report well structured and well reasoned?

Example Project

Example 1: Price forecast and virtual bidding of electricity contracts in the wholesale market

Great volatility exists in the wholesale electricity prices. With increasing penetration of renewable generation, such volatility could be magnified given the intermittent nature of the renewable energy sources. This project aims to develop efficient learning schemes that predict the real time electricity markets based on day-ahead information. Data can be collected from http://oasis.caiso.com/mrioasis/logon.do. Features such as historical day-ahead prices and load forecasts from PJM interconnection will be used for the prediction. The prediction would then be used for developing strategies of portfolio management in virtual bidding (see e.g. https://en.wikipedia.org/wiki/Virtual_bidding).

Example 2: Extraction road networks from satellite imagery

Recent flooding in Bangladesh and Hurricane Maria in Puerto Rico observe the need for accurate mapping during the disaster response. However, most existing map features such as roads, building footprints, and points of interest are primarily created through manual techniques. This project aims to develop advanced computer vision algorithms to process remote sensing data from https://spacenetchallenge.github.io/Competitions/competitionsSummary.html and automatically extract features to map the road networks from satellite imagery.

Example 3: Automatic music soundtrack generation

In order to augment the creative process, many tools now exist for helping artists write music. However, computer music generators typically operate at the note level and are able to capture long-term dependencies. In addition, they lack the expressive richness and nuance of performed music. This project aims to develop a novel deep learning model trained on the music dataset from http://www.piano-midi.de/ to compose music that reflect the nuance features such as pitch interval. timbre, tempo, rhythm.

Example 4: Fake Yelp review detection

Online reviews have become an important factor when people make purchase and business decisions. The increasing popularity of online reviews also stimulates the business of fake review writing, which refers to paid human writers producing deceptive reviews to influence readers' opinions. This project aims to building a classifier that takes the review text and the basic information of its reviewer as input (https://www.yelp.com/dataset/challenge) and outputs whether the review is reliable.

Notes:

- **Datasets:**
  - Kaggle
  - UCI Machine Learning Repository,
  - Amazon AWS Open Data
  - Google Dataset Search
- **Computing Resources:**
  - Northeastern Discovery: https://rc.northeastern.edu/
  - AWS Education Program: https://aws.amazon.com/education/awseducate/apply/
  - Google Colab: https://colab.research.google.com/