Machine Learning
CS 5350/6350, Fall 2015

home   information   topics   lectures   homeworks   projects   resources

Class Projects

(*This page will be updated to include more information about possible project ideas.*)

1. Overview
2. Milestones and grading
3. Project topics and ideas
4. Ideas

Overview

The goal of the final project is to show me (and everyone else) that you have learned something in the class. It is an opportunity for you to explore ideas that you see in the lectures and homeworks and extend them. You can even think of your project as first steps towards research in machine learning or its applications.

Projects should be done in groups of at most two students.

Milestones and Grading

Your project is worth 24% of the class grade. This is broken down across the following milestones:

1. **Form a project team** (5 points): Inform the instructor who is in the project group.
2. **Project proposal** (10 points): Submit a one page document that says what your project is. This document should contain the following information
   - Who is in the project group?
   - A very brief description of the problem you want to address.
   - Why is it interesting? How is it relevant to the class/machine learning?
3. **Intermediate status report** (20 points): A two page document that says what you have done on the project so far. By this time, you should have interacted with the instructor or the TAs about your project. The report should describe:
   - The progress you have made towards your goal. (This can not be just "We collected data".) (50%)
   - Details of your plan for rest of the semester (30%)

- Pointers to literature (20%)

4. **Final report** (65 points): The final report can be at most six pages (size 11 font). The report should be structured like a small research paper. Broadly speaking it should describe:

    1. What problem did you work on? Why is it interesting?
    2. What are the important ideas you explored?
    3. What ideas from the class did you use?
    4. What did you learn?
    5. Results (or for theoretical project, proofs)
    6. If you had much more time, how would you continue the project?

    Each of these components will be equally weighted in the report grade. *More details about the final report will follow*.

Topics

Any project with a significant machine learning component will be fine.

If you are looking for project ideas, come to the office hours and we can brainstorm ideas. Projects can be one of:

1. An application project. Eg: Some application of machine learning that you find interesting.

2. Replicating published results: Eg: You find a machine learning paper very interesting and want to reimplement their system to see if you can get the same results.

3. A theoretical project. Eg: Prove some interesting properties about a learning algorithm

4. An algorithmic project. Eg: Develop a new learning algorithm that applies to a particular kind of problem.

5. Your own research. Eg: If you are working on something already and wish to apply ideas from learning to it.

In general, pick topics that you find exciting. Try to convince me that the topic is interesting.

Important: Applications and experimental projects *should* follow rigorous experimental procedure (i.e. explore features, build fair baselines, use cross-validation for hyper-parameter selection for all settings, report both positive and negative results.)

Project Ideas

You are encouraged come up with your own ideas for projects. For example, you can look for data for a problem that you care about and try out different learning algorithms to predict

something interesting about the data. For application projects, your experiments need to be rigorous.

You may also look at the various current and past competitions on Kaggle for inspiration.

You are welcome to explore and expand upon project ideas listed below or use these as a starting point for brainstorming ideas. If you want to discuss these further, please come to my office hours.

Learning protocols, algorithms and theory:

- Why does a linear predictor make a prediction? Can a classifier make a prediction and provide a human-interpretable explanation for its prediction?
- Active learning: Exploring different kinds of interactions between the teacher and the learning algorithm. Can a learner learn from a weak teacher? Active learning with Mechanical Turk. Active learning to explore new features.
- Comparing different regularizers and loss functions, either experimentally (easier) or analytically (harder).
- Learning kernel functions or distance metrics
- Comparing different optimization algorithms either experimentally or analytically
- Exploring multi-task learning: Learning to perform multiple tasks together.

Applications:

- Art: What is the genre of a song/book/painting? Who is the artist/composer/writer/performer? Chord detection or classification.
- Learning to play board games like Risk, Settlers of Catan, etc.
- Playing the stock market: Will the stock market go up or down tomorrow? Should a stock be purchased? Can we predict stock trends using Twitter?
- Astronomy: Classifying astronomical objects (quasars vs. red giants, vs white dwarfs, etc); discovering extra-solar planets
- Shopping: Should I buy an item (eg. a flight ticket) now or wait?
- Biology and medicine: Identifying disease risks, binding sites, etc from gene data
- Software and security
  - Was this piece of code copied from somewhere?
  - Identifying people by their computer usage
  - Detecting malicious software/methods
  - Authenticating cellphone users using biometric or accelerometer signals
- Language, speech and vision: There are many applications of machine learning in these areas. Talk to me if you want other ideas.
  - Detecting sarcasm, humor, irony, metaphor, etc in text
  - Do two sentences or paragraphs mean the same thing?
  - Detecting pedestrians in street images or videos
  - Identifying events in videos. (maybe using both the video and subtitles)
  - Detecting plagiarism

- Character\face\object recognition or classification
- Finding specific voice in a noisy data

Remember: *This list is just meant to give you ideas and is not an exhaustive list of projects.*

home   information   topics   lectures   homeworks   projects   resources