

Final Project Guidelines

*Handed Out: March 14th, 2018**Due: April 27, 2018*

- You may work in teams of 2-3; larger teams will be expected to have more ambitious projects. Team members will evaluate each other and I expect reasonable load balancing of work.
- The final project report is due by **April 27th, 2018**. Your project should be ~7-10 pages in length, 11 point font, single-spaced. However, you can have as many pages of references as you would like. I will start reading them April 28th around 3p. Modulo extenuating circumstances, no late projects will be accepted once I start reading. This is worth 60% of the project grade.
- A draft final project report is due by **April 20th, 2018**. It is worth 5% of the project grade.
- The final presentation will be on **April 18th, 2018** during class. However, the emphasis here will be on the method used and therefore do not require final results. I expect a 15-20 minute presentation using slides. This will be worth 30% of the project grade.
- The initial proposal is due **March 28th, 2018** and worth 5% of the project grade. I will start reading them March 29th around 5p – please make this easy for me.
- Please submit your solutions via your CCIS `github` account. This includes a PDF file of the report, \LaTeX source, slides, and code. If you are submitting any code, instructions for running this code should be both in the email and the report (copy and paste is fine). ☺

Overview

Unlike most final projects that emphasize an application, I would like to balance the emphasis between a deeper dive into methods related to those covered in the course and/or a potentially interesting application. Accordingly, below are some topics we covered and potentially related methods. These are just ideas – you may also want to consider empirical or theoretical studies around variants of methods we discussed in class.

Topic	Advanced Topic
Decision Trees	Regression Trees
Generative Models	Bayesian Networks, Hidden Markov Models, Topic Models, Variational Methods
	Generalized Linear Models
Regression	Alternative Estimation Methods (e.g., dual coordinate ascent, etc.)
	Markov/Conditional Random Fields
Online Learning	Bandit Algorithms, Structured Perceptron
Kernel Methods	Scaling SVMs, Specialized Kernels, Structured SVM
Ensemble Methods	Gradient Boosting
Practical Issues	Advanced Evaluation Methods, Class Imbalance, Cost-sensitive Learning, Decision Theory, <i>many others</i>
Deep Learning	Representation Learning, <i>many things</i>

Other interesting topics would be reinforcement learning, alternate learning protocols (e.g., active learning, semi-supervised learning), alternate learning settings (e.g., domain adaptation, multi-task learning, transfer learning), dimensionality reduction. I'm fairly flexible in this regard – my only rules are that the rest of the class should be able to understand your presentation. Regarding the actual implementation component, I am open to anything between further investigation into an existing work (e.g., add more datasets and try to characterize findings not present in the paper) to a new dataset and application. We will refine this with the proposal.

Proposal

The purpose of the proposal is to provide sufficient information for me to comment and for us to agree on the expectations of your particularly project. For this, I would like a 1-2 page document consisting of

1. The method you are looking to study.
2. How this relates to topics covered in the course.
3. What you are specifically interested in studying.
4. Why you think this is an interesting problem (and why potential future job interviews might find it interesting). Ideally, this could eventually turn into a “real” research project.
5. Reference to any datasets that you plan on using (and if you already have them). Note that this may not be relevant for more theoretical studies.
6. Related work – I expect at least 2-3 citations.

Preliminary Report

Details to follow. Basically, I expect an abstract, section headers, partially written introduction and related work sections, properties of any datasets used, and placeholders for any empirical results with captions.

Presentation

Details to follow.

Below are some basic guidelines regarding the writeup for your final project. Please note that these are only guidelines – the form of your report should be tailored to the specific problem you are investigating. While your goal should be to highlight the most interesting aspects of the work you have completed (and therefore the structure is completely up to you), I would generally expect that the report follows a basic “IMRAD”-like structure. Furthermore, while i encourage creativity, I encourage you to make your report look like a publishable article. Specifically, I would like you to use the L^AT_EX style files from any major venue or publisher (e.g., ACL, ICML, NIPS, ACM, LNCS, etc.) – but this is not *technically* required. Anyway, below are aspects I will be looking for; however, please do *not* believe that you should use the headings below as the exact headings of your paper (although some will be the same...I assume).

1 Introduction

Motivate and contextualize the problem you are investigating.

- What is the basic problem?
- Why is this problem important?
- How does this work fit amongst related work in this area?

Provide a high level view of how you are addressing the problem and summarize your contribution.

- What is your approach?
- Why did you choose this approach?
- Summarize the basic results and conclusions in 1-3 sentences.

2 Problem Definition

2.1 Task Definition

Formally introduce the model and/or model you are investigating and define the notation you will be using throughout the report. Precisely specify the inputs and outputs of your problem. Furthermore, elaborate on why this is an interesting and important problem (particularly if you are doing a survey or “formal” proposal).

2.2 Algorithmic Specifications

If you are studying a specific learning algorithm or set of learning algorithms, this is the appropriate place to present this. Describe the algorithm(s) you will be using in sufficient detail. Pseudocode, figures, and/or motivating examples are frequently useful here.

2.3 Expectations

This is where you should clearly state your hypotheses. In the case of an experimental study, discuss what you hope to achieve and what you expected the results to be. How do you expect each algorithm to behave and why? Discuss why you chose this specific experimental design to investigate your problem of interest. If you are doing a “formal” proposal, this component would be particularly important.

3 Experimental Methods and Results

3.1 Experimental Methodology

What are the criteria you will be using to evaluate your method? Elaborate on the details of your experimental design. How does the training/test data used represent a *real-world* setting? What experimental results will be collecting and how will you be analyzing it?

3.2 Results and Analysis

Present the quantitative results of your experiments. Of course, I would really appreciate graphical presentations of data although tables are also appropriate in many settings. What are the basic results revealed by your experiments. Are the results statistically significant?

3.3 Discussion

Was your hypothesis supported? Compared to other methods, what were the strengths/weaknesses of your method? How can the results be explained in terms of properties of the algorithms and/or data? If you had infinite resources and/or information, how could you conduct more convincing experiments?

4 Theoretical Evaluation and Analysis

If your study is more of a theoretical contribution, survey, or expanded proposal – this would be the section for your analysis and contribution. Please make it clear what is derived from known work and what should be reconsidered by your analysis/presentation. What is novel from your perspective of the problem. I welcome position statements here – particularly if they are well supported by convincing arguments.

5 Related Work

One aspect of this project that I believe important is your knowledge of some of the related work (as I want you to read research papers – but specifically those you find interesting to your work). This is another opportunity to contextualize your work, but from a more contrastive perspective. How is your problem and/or method unique? Why is your problem and/or method better? Please be aware that I read more papers than just about anybody you know, so while I do not expect this to be exhaustive, I do expect you to be aware of the “major” works. ☺

6 Future Work

What are some of the significant shortcomings of your method? For each shortcoming, propose how you might be able to overcome them with more time and/or resources. Given what you have learned, what might you do differently?

7 Conclusions

Briefly summarize the important results and conclusions presented in the paper. What is the take away message of your work?

I will be evaluating your work along many dimensions including novelty, effort, clarity of presentation, completeness, interestingness,¹ etc. However, I *strongly* encourage you to make a *clear*, *interesting*, and *convincing* argument. Please be aware that if you turn in your project after the final, I will be spending 20-30 minutes reading each project.² Therefore, if you can make your project stand out (although I will not be grading on a curve with respect to the projects), I would be most appreciative.

¹And you thought patent law doesn't need reforming United States Patent Application 20060242139.

²Although I review tons of papers, so I am actually pretty good at this.