CS 475: Final Projects
Prof. Mark Dredze
Version 1.0

Important Dates:
March 24: Project proposals due
April 14: Progress reports due
May 12: Final project writeups due
May 14, 9-12pm: Poster session

All submission should be by email to cs475@cs.jhu.edu.

The class project is worth 25% of your final grade. You should plan on spending an appropriate amount of time on the project (30 hours per person, 60 hours per two person project.) Projects will be done in groups of 2 (or 1 if you write a tutorial). Both people on a project will receive the same grade, so you are responsible for choosing a responsible project partner.

Project Types
Machine Learning Systems
Implement and apply a machine learning algorithm to solve a problem. In doing so, you must demonstrate knowledge of something we've learned in class, or a general topic in machine learning. A simple black box application of machine learning is insufficient. You may also implement a machine learning algorithm based on a research paper of interest. You should explain how you will implement the algorithm and what new experiments and settings you will explore.

This is a good opportunity to work on a research project, especially if you are a PhD student. My general advice to PhD students is to work on something that can become a paper, although you probably won't get that far in the class project. Projects on learning theory are also acceptable. Both individual and teams of two can work on these projects. Most students choose this project.

Your proposal should identify which machine learning concepts and ideas you'll be testing, the data you will need for this project, and an expected outline for your final writeup. Data collection cannot be a significant part of the project, so be clear as to where the data will come from and what state it is in. Additionally, please include:
- Resources you will use (libraries, data, etc.)
- Methods: a description of the machine learning method you will use
- Milestones: the deliverables for the project. This should include must achieve, expected to achieve, and would like to achieve
- An outline of the final writeup. This is a description of what I can expect you to deliver in your final report.
- Bibliography- a list of relevant citations

Survey/Tutorial
Write a comprehensive tutorial or survey on a specific topic in the literature. Your paper will be about 25 pages and must reflect 30 hours of work. The quality should be such that you can post the tutorial online for others to use. You should explain the topic in depth, and not simply create a list of explained references.

Your proposal should describe the topic, what areas you will focus on, and include several references you will be using. You should include a proposed outline of the tutorial. Specifically, you should include:
- An outline of what you will include in your tutorial/survey. This should include must achieve, expected to achieve and would like to achieve (reach goals)
- Bibliography: a list of papers that will be included in your survey. This should include a brief description (1-2 sentences) of each paper and how it fits into your tutorial.

Project Proposals
Your project proposal should be at most one page. A single proposal is needed for each team and will be submitted by email. Your proposal should include your name, email, the type of project you are doing, and the details listed above. **Please complete the provided project proposal template.**

Progress Reports
The progress reports should be short (1 paragraph is sufficient in most cases.) One report is needed per group and will by email. You do not need to create a PDF. Your progress report should:
- briefly summarize what you've done so far
- compare to your proposal: are you on track, a bit off track, major changes needed
- describe any changes you anticipate in your final report from your proposal

Final Project Reports
The following is a guide for the implementation projects. If you are unsure of what to include in a survey, please email me.

The final report should be a SIX page PDF (not including references). This is not a requirement but a strong suggestion. Slightly longer or shorter writeups are acceptable, and you may include as many appendices as needed. The format of your report should use the NAACL 2010 format, available in both Latex and Microsoft Word: http://naaclhlt2010.isi.edu/authors.html

Please be sure to include your names in the report.

Your report should follow your proposed outline. In addition, it should include the following:

- An introduction and background section that explains relevant material and situates your work in the context of existing literature. Do not assume the reader knows the details of your field. You must include sufficient references.
- An explanation of all machine learning techniques used as well as an explanation of why they were used. This should reflect the knowledge you've learned by taking this course. Don't just say "I used an SVM" but explain what an SVM is in detail. Remember, this report is for a machine learning class, and that should be the focus of the writeup.
- A detailed description of the work done for the project. Note that in a conference or journal publication, this section is focused on the science and not the details of the actual work. For these writeups the focus is shifted. I want to know what you did, i.e., how you prepared data, the libraries you used, code you wrote, problems you had, etc. This is where you explain exactly how you spent your time on the project. This should convey exactly how much you did.
- A description of the results you obtained. Examples of actual output (pictures, sentences, etc.) would be great if appropriate. Please include details for any tests you ran, even if they weren't "interesting" or didn't work. Again, this should highlight the work you did. You should include an analysis of the results using knowledge of machine learning (I think this worked or didn't work because this algorithm assumes...)
- Comparison to proposal. This should include a detailed list (bulleted lists are appropriate) of your progress measured against your proposal. This part is similar to your progress report in that it includes a list of what you were able to accomplish, changes to your original plans, etc.

I found the following writeup advice from Tommi Jaakkola for his 2004 machine learning class. I think it's very relevant so I'll just quote it:
"We expect that the ``size'' of your project should be equal to about the amount of work required for ...(Mark: 25% of your grade). The project, however, should be in some sense ``complete''. By this we mean that you should not ignore relevant machine learning issues. In the final report you shouldn't just say what you did but also why it was a reasonable thing to do given the course material... You shouldn't worry about getting ``great'' results. The idea and your understanding of the machine learning issues involved are much more important than getting ``great'' results."

Poster Presentation
Details on posters will follow.

Grading
The project is worth 25% of your grade You are expected to spend an equivalent of hours per person on the project (about ~2.5 homework assignments). I will be evaluating your project based on the writeup using the following guidelines.

Did the project constitute a sufficient amount of work for a final report?
Is the writeup clear and understandable?

Does the writeup demonstrate an understanding of machine learning and are relevant machine learning concepts and algorithms explained?
Does the writeup include sufficient references and descriptions of related work?
Was progress made along the proposed goals and if not, is there sufficient explanation given?
Does the writeup clearly present the performed work?
Are the results analyzed and understood using machine learning concepts?

Code
If you wrote code, please submit it along with the writeup. You do not need to include directions for running the code nor do I expect it to work if it needs external libraries, data, etc.