CS 229, Autumn 2007
Final Project Guidelines and Suggestions

1   Project overview

One of CS229's goals is to prepare you to (i) apply state-of-the-art machine learning algorithms to an application, and (ii) do research in machine learning. The class's final project will offer you an opportunity to do exactly this.

The important dates for the CS229 project are:

- Proposals: Due at noon on Friday, 10/19.

- Milestone: Due at noon on Friday, 11/16.

- Poster presentations: Morning of Wednesday, 12/12.

- Final writeup: Due at 11:59pm on Friday, 12/14 (no late days).

Projects can be done in teams of up to three students. If you have a project of such grandiose scope and ambition that it cannot be done by a team of only three persons, you can propose doing a project in a team of four.

2   Project topics

Your first task is to pick a project topic. If you're looking for project ideas, please come to either Prof. Ng or the TAs' office hours, and we'd be happy to brainstorm and suggest some project ideas. In the meantime, here are some suggestions that might also help.

Most students do one of three kinds of projects:

1. **Application project.** This is by far the most common: Pick an application that interests you, and explore how best to apply learning algorithms to solve it.

2. **Algorithmic project.** Pick a problem or family of problems, and develop a new learning algorithm, or a novel variant of an existing algorithm, to solve it.

3. **Theoretical project.** Prove some interesting/non-trivial properties of a new or an existing learning algorithm. (This is often quite difficult, and so very few, if any, projects will try to do this.)

Some projects will also combine elements of applications and algorithms and theory.

Many fantastic class projects come from students picking either an application that they're interested in, or picking some sub-field of machine learning that they want to explore more, and working on that as their project. If you haven't worked on a research project before but would like to, you can also use this as an opportunity to try your hand at it. (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 machine learning might be applicable to, then working out how to apply learning to it will often make a very good project topic. Similarly, if you currently work in industry and have an application on which machine learning might help, that could also make a great project.

A very good CS229 project will comprise a publishable (or nearly-publishable) piece of work. Each year, some number of students continue working on their projects after completing CS229, and submit their work to a conference or journal.

So, for inspiration, you might also look at some recent machine learning research papers. Two of the main machine learning conferences are ICML and NIPS. You can also find papers from recent ICML conferences online (`http://www.icml2006.org/icml2006/technical/accepted.html`, `http://oregonstate.edu/conferences/icml2007/paperlist.html`). All NIPS papers are online, at `http://books.nips.cc/`. Finally, to see a list of last year's class projects, you can go to `http://www.stanford.edu/class/cs229/projects2006.html` .

Projects will be evaluated based on:[1]

- The technical quality of the work. (I.e., 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?)

- Significance. (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 impact?)

- The novelty of the work, and the clarity of the writeup.

Lastly, a few words of advice: Many of the best class projects come from students working on topics that they're excited about. So, pick something that you personally can get excited and passionate about! In addition, don't be timid, and when in doubt go for whatever's the more ambitious option. Finally, if you're not sure what would or would not make a good project, please feel **strongly** encouraged to either email us or come to office hours to talk about project ideas.

3   Project submission logistics

This section contains the detailed instructions for submitting different parts of your project. You probably do not need to read any of this in great detail until nearly the due date of the submissions.

3.1   Project proposals

Your proposal should be a normal (plain ASCII) email, giving the title of the project, the full names of all of your team members, and about a 300-500 word description of what you plan to do. Please send your proposal as a normal email and not as an attachment, or use any other document format (such as PDF or MS-Word).

---

[1]Don't overthink these criteria, nor worry too much if you're not sure that you can do well on all of them. Instead, just think of them as an "ideal" that you should aspire to, particularly if your goal is to do publishable research work.

3.2 Milestone

The project milestone is due at noon on Friday, 11/16, which is roughly halfway between the proposal and the final project due dates. Your milestone report should describe what you've accomplished so far, and very briefly say what else you plan to do.

The milestone will help you make sure you're on-track. You should write it as if it's an "early draft" of what will turn into your final project. Specifically, you can write it as if you're writing the first few pages of your final project report, so that you can re-use most of the milestone text in your final report. Please write the milestone (and final report) keeping in mind that the intended audience is Prof. Ng and the TAs. Thus, for example, you should *not* spend two pages explaining what logistic regression is.

Submission instructions: Your milestone report should be at most 3 pages long. Please send the milestone as an email attachment, to   . **Please submit your milestone in PDF format.** In particular, we do not accept MS-Word, OpenOffice, PostScript, or any other document format. You should name your PDF file according to the format "`yourConcatenatedLastNames-ProjectTitle.pdf`". For example, if the project partners are John Doe and Jane Smith, and your project title is "Learning to Recognize People," then name your PDF file `DoeSmith-LearningToRecognizePeople.pdf`.

In addition, in the body of the email (which should be plain ASCII, and *not* HTML), list the full names of all the team members on the first line, and state the full title of your project on the second line. (The rest of the email can be anything you like.) For example,

```
From: Jane.Smith@stanford.edu
To:
Cc: John.Doe@stanford.edu
Subject: Project milestone

John Doe and Jane Smith
Learning to recognize people
```

Please follow these submission instructions exactly. Failure to do so may result in our email system not receiving your file.

3.3 Poster presentations

The class projects will be presented at a poster presentation on Wednesday, 12/12. Each team should prepare a poster, and be prepared to give a very short explanation, in front of the poster, about their work. At the poster session, you'll also have an opportunity to see what everyone else did for their projects. (SCPD students living outside the bay area are exempt from this.) We'll supply poster-boards and easels for displaying the posters.

3.4 Final writeup

Final project writeups are due at 11:59pm on Friday, 12/14. **Late days cannot be used for the final writeup.** Final project writeups can be at most 5 pages long. Apart from the page limit, please follow the same submission instructions (such as filename and format of email) as the milestone.

If you did this work in collaboration with someone else, or if someone else (such as another professor) had advised you on this work, your writeup must fully acknowledge their contributions.

After the class, we will post all the final writeups online so that you can read about each others' work. If you do not want your writeup to be posted online, then please let us know at least a week in advance of the final writeup submission deadline, and we'll give you a different email address to which you may send your writeup.

4 Miscellany

If, after CS229, you want to submit your work to a machine learning conference, the ICML 2008 deadline will be in early February next year (see `http://icml2008.cs.helsinki.fi/` for details), and the NIPS deadline is usually in June (`http://www.nips.cc/`). Of course, depending on the topic of your project, other non-machine learning conferences may also be more appropriate.