CSI 5325 Project Guidelines

Greg Hamerly [*]

January 14, 2014

1   Project overview

The goal of the semester project is to help you work on a larger project related to machine learning. You will gain experience in the development, implementation, analysis, and/or application of the things we discuss in class. It will help prepare you for further work in machine learning, or help you use machine learning for another project.

The important dates for the project are as follows:

- Proposals: due via email by 5 PM on Tuesday, February 4.

- Milestone: due via email by 5 PM on Tuesday, April 1.

- 10-minute in-class presentation: during lecture of final week of class.

- Final writeup: due via email by 11:59 PM on Monday, May 5.

2   Project topics

You first need to pick a project topic. You can talk to professor Hamerly during his office hours about choosing a topic, and brainstorm with each other. There are typically three kinds of projects:

1. Application project. This is the most common type of project: pick an application that interests you, and explore the best way to apply machine learning to solve the problem.

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 or non-trivial properties of a new or existing learning algorithm. This is typically quite difficult.

Great projects can come from students combining their interest in an application with things they're learning from this class. So it's good to choose something you're excited about. This is a good chance to start work on a research project.

Getting started and choosing a topic can be a bit difficult, so it's good to look around for other ideas as inspiration. Good places are talking with your classmates or your professor, and looking at published work in machine learning conferences such as ICML or NIPS (the papers for these conferences can be found online; see `http://machinelearning.org/icml.html` and `http://books.nips.cc/`). A good project writeup will follow the style of these papers, and be of publishable quality.

Projects will be evaluated based on:

---

[*]Based on Andrew Ng's CS 229 project guidelines, used by permission.

1. The technical quality of the work. (That is, do the technical choices make sense? Is the approach reasonable? Are the proposed algorithms or applications clever and interesting? Do the authors convey novel insight about the problem and/or algorithms?)

2. 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?)

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

3 Project submission details

Here are more details on submitting the different parts of the project. Please see the important dates at the beginning of this document for when each part is due.

3.1 Project proposals

Project proposals should be emailed to Dr. Hamerly. Your proposal should be a plain ASCII email giving the title of the project, and a 300-500 word description of what you plan to do. Please send your proposal as a normal email and not as an attachment.

3.2 Milestone

This report should describe what you've accomplished so far, and very briefly state what else you plan to do.

The milestone will help you keep on track. You should view it as an early draft of the writeup you will turn in at the end of the semester. Specifically, you can write it as if you're writing the first few pages of your final report, so you can 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 Dr. Hamerly or others familiar with machine learning. Thus, for example, you should *not* spend two pages explaining what logistic regression is.

Your milestone report should be at most 3 pages long. Please send the milestone as an email attachment to Dr. Hamerly. Please submit your milestone in PDF format, using LaTeX to prepare the document. The name of the PDF should be formatted as "`lastname-ProjectTitle.pdf`", where lastname is your last name, and ProjectTitle is the abbreviated title of your project.

As you write the milestone and final report, please pay attention to follow a similar format of other machine learning research papers (see ICML and NIPS conference papers above). In particular, many experimental papers have the following structure:

- **Abstract** – A brief overview of the paper.

- **Introduction** – Describes the motivation of this work and outlines the rest of the paper.

- **Background and/or Related Work** – Describes what other researchers in the same area have done, and how they perhaps could be improved.

- **Methodology** – Describes what is the approach taken in this paper.

- **Experiments** – Describes the experiments performed, including details on the data used.

- **Discussion and/or Analysis** – Examines the results of the experiments and draws some conclusions about their significance.

- **Conclusion** – Summarizes the paper.

- **References / Bibliography** – Gives properly formatted references to other scholarly work that this work is built on. Note that the references should be scholarly, which means things like refereed conference and journal articles. Importantly, that rules out things like most websites, basic textbooks, and press articles.

3.3 In-class presentation

You will give a presentation in class on your work during the last two lectures of the semester. Each person is expected to give a 10-minute presentation with powerpoint-style slides that describes your work. Ten minutes is not much time, so keep it brief – at most 15 slides total. We will adhere to the 10 minute schedule so that everyone has a chance to present, and so that there is time for questions after each talk.

3.4 Final writeup

Final project writeups can be at most 8 pages long. Please name the PDF in the same way as for the milestone.