COMPSCI 271 - Final Project

Due date: Tuesday Dec. 13, 11:59:59

1 Project Scope

For this project you and a partner will evaluate a machine learning technique theoretically, or on some benchmark or real world data. Acceptable projects include the following:

1. Proving a new theorem about an existing learning method (ambitious).

2. Proposing a modification to an existing learning method and evaluating this modification against the standard method on several benchmark data sets.

3. Developing a hypothesis about how different learning methods compare and evaluating this hypothesis using these learning methods on many different benchmark data sets.

4. Proposing an application of machine learning to an interesting problem that you care about, selecting methods to try, doing your best to make them work, and discussing the results.

This is only a partial list. Other ideas are also possible if you clear them with me first.

I expect that most or all projects will be two person projects. If you have a good reason why you should do a project with a different number of people, please contact me ASAP.

You should aim to do something novel in your proposal. It does not need to be a publishable result, but you should aim higher than merely reproducing a known result. There can be novelty in your choice of techniques and/or the match between techniques and problems.

2 Project Topics

I expect that most people will wind up applying machine learning to some topic that is of personal interest and/or that is related to ongoing Master’s or Ph.D. research. This project is a great opportunity for you to explore machine learning in the context of a problem that matters to you or to add a new angle or twist to existing research projects. I would be delighted if your machine learning project became a research paper or a chapter in your dissertation.

If you are are struggling to come up with project ideas, I can suggest a few sources of inspiration:

- The kaggle site (linked from the resources section of the class web page) has information about machine learning competitions.

- Representatives from the GoodFit startup have suggested some project ideas. I can provide extra copies of their materials or put you in touch with their staff if you are interested.

- I have a few project ideas using reinforcement learning that I would like to see somebody try out at some point. One involves using a simulation of HIV to improve drug treatment plans.
3 Project Proposal

You should turn in a proposal for your project to me (email is fine) *no later than November 10*. This should be an approximately one page description of what you plan to do. You are, of course, free to (and encouraged to!) turn in a proposal earlier. Your proposal should include the following:

- The names of both people on the team.
- The goal of the project.
- What you expect to learn or discover from doing the project.
- At least 3 references that are relevant to your project. This is to spare you the embarrassment of not having an answer when I ask if you’ve done a google search on the keywords for your project.

I will give feedback on your writeup within a week of receiving it, but do not wait to receive feedback. You should continue to work on your project while I am reviewing the proposals. I’ll try to flag proposals that are in greatest need of attention and give feedback sooner rather than later.

4 Project Milestone

You should turn in an update on your progress on your project since your proposal by email *no later than November 29*. Your update should include new things you have done and any steps taken to address issues that I raised in my response to your proposal.

5 Project Write Up

By *December 13*, you should turn in a document describing your efforts and a link to any code you’ve produced. Note that it is perfectly fine to use routines built-in to Matlab or that you have found in reputable software distributions *if* you cite your sources.

As a rough guideline, you should aim for a writeup of 10-15 single column, double spaced pages. However, I won’t be counting pages and this is only a very rough guideline. You shouldn’t pad your writeup with worthless text to reach this size. If you can describe what you’ve done clearly and concisely in much less space than this, that’s great. I’ll expect your writeup to address (at least) the following issues:

1. What problem does your project address? If you have chosen an application area, please remember that I may not be an expert in the application you have chosen, so be sure to describe the application area clearly.

2. What methods did you use to address the problem?

3. What is the reason you picked the methods you picked? Can you justify theoretically or empirically that this was the best choice?

4. How did the machine learning techniques that you applied perform?

5. How did you validate your results, i.e., what were the training and test sets?

6. What difficulties did you encounter and how did you try to overcome them?

7. What would be the next step if you were to extend this project?

8. What did you learn from this?
Remember to include a complete bibliography.

Note that it is not a requirement for a good project to have successful experimental results. Sometimes things don’t work out as you planned. Your grade will be based upon your scholarly and insightful application of the things you have learned in class.

Your final grade on the project will be determined by the following factors:

- Proposal: 10%
- Milestone: 15%
- Final writeup: 75%

The only way to get less than full credit for the proposal is to fail to turn in something remotely reasonable on time. One can fail to get full credit on the milestone by failing to turn it in on time, failing to respond to feedback given to the proposal, or failing to demonstrate any progress between the proposal and the milestone.