Guidelines for the Course Project
CSG220, Spring 2007

Important Dates

Written Project Proposals Due: March 1.
Written Project Reports and Oral Presentations Due: April 26.

General Information

As examples of what you may do for your project, you may:

- apply a machine learning algorithm to an interesting task;
- implement two or more approaches to the same task and compare them;
- experiment with a novel and potentially worthwhile modification of an existing machine learning algorithm;
- experiment with an approach that combines two or more existing machine learning algorithms in a novel and interesting way; or
- perform a relevant theoretical analysis.

Resources for Possible Projects

A list of titles of projects from past versions of this course is available from the course web page. Some of the actual written reports are also available to be looked at individually, as explained in that document, while others are not.

Also, if you follow the link to the textbook web page, you will find links there to materials for two projects discussed in the textbook: backpropagation for classifying face images and naive Bayes for classifying newsgroup postings. The materials for each consist of programs and data files. Either of these is acceptable as the basis for your project, but I will expect you to go beyond those canned projects in some way.

You may find it desirable to use existing software for your project, adding your own modifications only as necessary. Note that there is a whole suite of freely accessible data mining routines (in Java) at the Weka web site (but no neural network or SVM software, as far as I know). Feel free to use what's there if it's suitable for what you want to do. If you want to use support vector machines, you should download the free SVMlight software (written in C) from the appropriate site. Web searches are bound to turn up other freeware that may be appropriate for what you want to do. As the course proceeds I will bring sites containing other useful software to your attention and/or place links to them on our course web page. (The software provided by me under the "Homegrown Machine Learning Software" link may not be suitable for "industrial-strength" applications, but you are certainly welcome to modify it in appropriate ways for this project if you wish.)

You may also find the UCI machine learning data repository useful. (See the link on the course home page.) It contains a large number of datasets that many machine learning and data mining researchers have experimented with.

Details of Project Proposals

Your proposal should describe the nature of the work you intend to do. The proposal should be no longer than 2 pages. If you intend to study a particular application area, be sure to describe both the nature of the application and the type(s) of learning algorithm you expect to use. Be reasonably specific.

For example, don't just say you intend to use such-and-such a (supervised) learning algorithm in such-and-such an application, but specify more precisely exactly what input-output mapping is to be learned, what the training data will consist of, etc. Also be sure to indicate what kind of representation you expect to use for inputs and outputs, etc.

When I give you feedback on your proposal at the following class meeting, I may ask you for more details on these issues if I think your proposal is too vague on any of these points. Of course it is okay if part of the work you propose involves exploring different learning algorithms and/or different data representations in a given application; just be sure to say so in your proposal (ideally, listing the possible alternatives you intend to explore).

Details of Project Reports

Your project report should include the following:

- problem definition;
- related work (including at least 2 literature citations and discussion of their content and relevance to your project);
- your approach;
- experimental results (unless your project is entirely theoretical);
- conclusions and ideas for future work;
- list of references; and
- appendices (if appropriate).

If you write any software yourself specifically for this project (or make significant modifications to anyone else's software), include the listings (of only the parts you wrote or modified) in an appendix. If you use software or data from some other source, be sure to cite that source (including the URL if obtained from the web), but do not include program listings for it.