Due: Various Dates
Marks: 40% of final grade. Late Policy: **0 Points After 12/15/04, 20% for each partial/full week after deadlines**

Policy on Plagiarism and collaboration: The project contents must be completed by only yourself.

The term project is a mini-research project. Machine learning is a constantly developing area of A.I and hence there are many open questions. A good term project will carefully select question/issue that can be addressed within a 10 week period and clearly and concisely answer the question. To help you achieve this, there will be many deliverables you must maintain along the way to the December 15th deadline.

The term project **must** have a distinct and non-trivial machine learning focus. Argue in your proposal why your chosen topic is a worth exploring, and that you are capable of addressing it. You will design and implement a solution to the problem you have chosen, and measure the performance of your solution. You will then write up a preliminary report which I will review and you will then turn into a final report. You will give a 20-minute presentation of the results of your project in class. This is not a software engineering course and hence though implementation of your solution is an essential component there are equally important components, such as the evaluation, presentation and discussion of your results. You need to regularly update me on how the project is progressing every two weeks.

Common problems in previously years have been:

- Starting excessively late.
- Going into indepth detail about implementation (this is not 518). Think about the project description as a submission to a conference (in fact previous projects have been published in conferences)
- Not putting in enough effort into the proposal (note its worth a lot but the write up is short)
- Underestimating the time to write up the project. If I can’t understand it, it doesn’t count.
- You did not demonstrate your approach addressed the problem.
- Poor writing style, this is a computer science course and there is no room for verbose prose or rhyming sentences.

For the term project you may choose your own topic (if you’re a Ph.D. student in good standing). If you choose this option, then you are responsible for all aspects of the project. Otherwise choose a topic from the list of areas on the following page. If your background is in another discipline you are encouraged to pick a topic that involves the application of machine learning to a solve a problem of interest in your area of expertise.
For any topic covered in class you can propose a term project centered around it. What follows
Below are additional topics not covered in the class.

Applications

Face Recognition
(see Web page for this course in 2003)

Handwriting Author Recognition
(see Web page for this course in 2002)

RoboCup Simulator
(see Web page for 535 course in 2004)

Information Extraction (Jointly with Prof. Haas SUNY – Albany)

“I have a corpus of simple English sentences tagged with semantic interpretations. The sentences are directions -- things like "Go down to the end and turn left. It's the 3rd door on your right". The semantic interpretations are simple slot-and-filer structures: "last door on left" is an object with type=door, direction=left, number = last. Every filled slot is marked with the number of the word that filled it.”

From this data you will need to learn how to fill in the slots for new test instances.

Theory

Ensembles of Learning Techniques (Jointly with Wei Fan, IBM Watson Research Labs)


Learning Predictive State Representation (Jointly with Csaba Szepesvári, Hungarian Mathematical Institute)


There are many intermediate steps in this project. You need to follow all of them to stay on track. The steps are:

10/08/04: Send email to davidson@cs.albany.edu nominating the area you will focus on.
10/21/04: Send me your project proposal (5%) (2 page)

I suggest the following format, from which, of course, you may deviate.

- [Abstract] The essential aspects of your proposal.
- [Problem Statement] A clear statement of the problem to be solved.
- [Motivation] Some background and motivation indicating that this is a problem worth solving.
- [Hypothesis] The solution you propose to the problem, with some preliminary justification.

I'll respond around 10/28/04

11/10/04: Extend your proposal to include a plan (7.5%) (3 pages)

I suggest the following format, from which, of course, you may deviate to suit your particular needs.

- [Method] The method by which the solution is going to be evaluated. Be specific: state which measurements you intend to make, what you expect to learn from them, etc.
- [Expected Results] Be specific about the results you expect to obtain. If you have some preliminary results, state them. If you are doing an experiment, try to make an educated guess at what the results will be. For instance, if you are going to measure average time to solve instruction scheduling problems, it is much better to say “We estimate solving a benchmark instruction scheduling problem will require 3000 random restarts of GSAT”, than to say “We will report the results of measuring times for solving instruction scheduling problems”. Of course, nobody is going to hold you to your estimates. Nevertheless, discrepancies between estimates and results are often the source of an interesting discussion.
- [Related Work] Relate other work to what you have done.
- [Timetable] Give some convincing arguments why you believe you are able to carry out the project in the remainder of this semester.

12/07/04  Student Presentations (7.5%)
12/09/04
12/14/04

12/15/04  Final Report (20%)

The final report should follow essentially the same format as the proposal. The problem statement and motivation sections would typically be grouped into a single introduction section. The section on expected results should be replaced by a section containing the results you obtained, with some discussion. The timetable should be deleted and replaced by a section containing conclusions and avenues for further work.

The report should be professionally written up with no spelling mistakes, a full set of references and be clear and succinct.