

# ComS 573 – Machine Learning

Vasant Honavar  
Artificial Intelligence Research Laboratory  
Department of Computer Science  
226 Atanasoff Hall  
Iowa State University  
Ames, Iowa 50011

Last updated: January 10, 2003.

## 1 Course Objectives

The primary objective of this course is to provide an introduction to the basic principles, techniques, and applications of Machine Learning. Programming assignments are used to help clarify basic concepts. The emphasis of the course is on teaching the *fundamentals*, and not on providing a mastery of specific commercially available software tools or programming environments.

In short, this course is about the principles, design and implementation of *learning agents* — programs that improve their performance on some set of tasks with experience. Upon successful completion of the course, you will have an understanding of a variety of practically useful machine learning algorithms and their use in data-driven knowledge discovery and program synthesis. You will have designed and implemented several machine learning algorithms in Java. You should also be able to adapt or combine some of the key elements of existing machine learning algorithms to design new algorithms as needed. The emphasis of the course is on machine learning algorithms that are useful in practice. However, we will cover some of the key theoretical results that have significant practical implications.

## 2 Prerequisites

The prerequisites necessary for fully benefiting from the material covered in this course include knowledge of data structures (e.g., lists, trees), design and analysis of algorithms, programming language concepts (e.g., functional programming, object-oriented programming, recursion, abstract data types), and selected topics in mathematics (e.g., boolean algebra, set theory, probability theory, calculus). If you are not sure whether you have the necessary background, please talk to the instructor.

Students will be expected to familiarize themselves with Java on their own, with the help of online resources and lab assignments.

In light of the emergence of Java as an attractive, platform-independent language for object-oriented design and implementation of software systems in general, and distributed artificial intelligence systems in particular, laboratory assignments will require you to program in Java. If you do not know Java already, you are expected to quickly acquire a working knowledge of Java on your own. Java should be an easy language to learn for those students who have a good grasp of object-oriented programming in a modern high-level language.

### 3 Course Staff

The instructor for the course is Dr. Vasant Honavar ([honavar@cs.iastate.edu](mailto:honavar@cs.iastate.edu)). The teaching assistant for the course is Mr. Jun Zhang ([jzhang@cs.iastate.edu](mailto:jzhang@cs.iastate.edu)). Additional information (office hours, etc.), will be posted on the course Web page at: <http://www.cs.iastate.edu/~cs573x/>.

The instructor and the TA will be available to answer your questions during the scheduled office hours, or at a time arranged by prior appointment, and at other times if necessary (if our schedules permit). The course Web page will be used to convey or update information concerning homework assignments, post lecture outlines, etc.

We might also use electronic mail to reach you when necessary. You are therefore strongly encouraged to get into the habit of reading your electronic mail and checking the course Web page once a day.

### 4 Computer Accounts

We will be using the computing facilities of the Department of Computer Science for all course-related assignments. If you do not already have a login on the departmental computer systems, please make sure that you have one by the end of first week of classes.

The Computation Center and the Computer Science Department hold tutorials that are designed to help a new users to get familiar with their facilities. Please contact them for a schedule and information on signing up for one of these tutorials.

The course web page located at will be used to post course materials including assignments, study guides, etc.

As a student in a course offered through the Computer Science department and a user of the ISU computer facilities, you are to abide by the department's Code of Computer Ethics a copy of which will be provided to you. *Please note that any suspected violations of the code of ethics are viewed extremely seriously by the Computer Science department and treated in accordance with the university's policies on academic misconduct.*

### 5 Assignments, Examinations, and Grading

There will be regularly scheduled written assignments, assignments, examinations to help you learn the material and to help us evaluate your progress. You will also need to complete a term project.

There will be two examinations — given approximately around the 8th and the 15th week of the semester respectively. The examinations may have a take-home component. Written assignments will be handed out roughly every two to three weeks. Laboratory assignments will be assigned every two or three weeks. You should expect roughly 4-6 written assignments and about 4 laboratory assignments. Some of the written assignments will require you to read, understand, and critique current research papers that are related to the topics being covered in class.

The term project has to be a small research or design project cuminating in a written report and possibly a brief oral presentation. You may choose to work in small groups

(consisting of 2-3 members each) on the project.

Students should be actively thinking about potential topics for projects or term papers right from the beginning of the semester. A list of suggested topics as well as guidelines for the preparation of project reports and term papers will be made available in due course. The instructor and the TA will be available for consultation and guidance on the projects or papers as needed.

The course grades will be based on written assignments, (19%), laboratory assignments (19%), two examinations (19% each), term project (19%), and and participation in discussions in class (5%).

## 6 Policy on Collaboration, Late Assignments, Etc.

The primary purpose of the assignments is to clarify and enhance the understanding of the concepts covered in the lectures. Past experience with this course has shown that this is helped by increased interaction among students. Discussion of *general concepts and questions* concerning the problem sets and laboratory assignments among students is encouraged. However, each student is expected to work on the solutions individually. Sharing of solutions (including segments of code) to assignments is forbidden unless explicitly instructed otherwise. If you are unclear about this policy, please talk to the instructor *before* you act. **Suspected cases of academic misconduct will be pursued fully in accordance with ISU policies.**

*On late assignments, there is a late penalty of 5% of the grade per day up to a maximum of 7 days from the specified due date. Assignments that are turned in later than 7 days after the due date will be assigned zero credit. Rare exceptions to this policy might be made (at the discretion of the course staff) under demonstrably extenuating circumstances.*

## 7 Syllabus

The following gives a tentative list of topics not necessarily in the order in which they will be covered in the course: Algorithmic models of learning. Design, analysis, implementation, and applications of learning algorithms. Learning of concepts, classification rules, functions, relations, grammars, value functions, models, skills, behaviors, and programs. Agents that learn from observation, examples, instruction, induction, deduction, reinforcement, and interaction. Computational learning theory. Data mining and knowledge discovery using artificial neural networks, decision trees, bayesian learning, association rules, genetic algorithms, dimensionality reduction, feature selection, and visualization. Learning from heterogeneous, distributed, dynamic data and knowledge sources. Learning in multi-agent systems. Selected applications in automated knowledge acquisition, pattern recognition, program synthesis, bioinformatics, and internet-based information systems.

More detailed list of topics will be posted on the course web page. Specific reading assignments and brief lecture outlines will be placed on the course homepage periodically.

## 8 Textbooks, Lecture Notes, and References

The primary textbook for this course is *Machine Learning* by Tom Mitchell.

Lectures will often draw upon a variety of sources to supplement the treatment of topics available in the primary textbook. A number of books and references are available on reserve in the Parks library. A list of useful references will be posted on the course web page. Additional sources of useful material include major journals and conference proceedings in these areas. You are strongly encouraged to explore various machine learning resources on the World Wide Web (Check the course Web page for pointers).