Syllabus

Dates and Time: MoWe 5:20PM - 6:40PM
Room Location: S B UNION 237 WESTCAMPUS
Instructor: Luis Ortiz
Office: 2313-C Computer Science Building
Phone: (631) 632-1805
Email: leortiz@cs.sunysb.edu
Web Page: http://www.cs.sunysb.edu/~leortiz
Office Hours: Tue-Thu 2:20pm-3:50pm; or, by appointment

General Information (Undergraduate Bulletin)

Course Description: A course on the fundamental concepts behind intelligent systems that autonomously learn to perform a task and improve with experience. The course will cover learning frameworks and problem formulations, standard models, methods, computational tools, algorithms and modern techniques, as well as methodologies to evaluate learning ability and to automatically select optimal models. Simple applications to areas such as computer vision (e.g., character and digit recognition), natural-language processing (e.g., spam filtering) and robotics (e.g., navigating complex environments) will motivate the coursework and material.

Prerequisites: CSE or ISE major
Credits: 3
Grading: Undergraduate Graded
Fulfilling Program Requirements: May be repeated as the topic changes, but cannot be used more than twice to satisfy CSE major requirements.

Overview

Roughly speaking, machine learning techniques strive to automatically acquire expertise to effectively perform a task of interest by efficiently processing task-related information, such as a (usually large) data set of examples, and successfully extracting and generalizing knowledge embedded within the available information.

There is growing demand for computer scientists with proficiency in machine learning. For example, the advent of technology for the collection of vast amounts of digital data, such as that generated by an ever expanding population of Internet users, has increased interest in the development of machine learning software applications. Machine-learning-based technology such as driver assistance and voice-activated systems in cars, automatic system personalization and adaptation
to individual user preferences and behavior, speech-driven phone systems for customer service, speech-to-text capabilities, recommender systems and e-mail spam filtering are now commonplace.

Machine learning application areas include marketing, e-commerce, software systems, networking, telecommunications, banking, finance, economics, social science, computer vision, speech recognition, natural-language processing, and robotics. Some problems addressed using machine learning techniques include pattern recognition and classification, knowledge discovery and data mining, anomaly detection, credit/loan approval, credit-card fraud, quantitative trading, automatic categorization of very large collections such as web pages, documents and images, effective ranking of web search results (e.g., Google’s PageRank), face recognition, tracking, machine translation, and more recently, intelligent, adaptive control of virtual player behavior in video games, smart debugging of computer programs and memory management in operating systems. There is also recent interest in creating computationally tractable machine learning tools for recognizing and predicting general trends in individual or group behavior in large populations, such as spending behavior, adoption of new products, technology or habits, sharing in peer-to-peer systems, and predicting the development of online communities within large social networks such as Facebook.

Given the broad applicability of machine learning techniques, it is natural to expect the need for computer scientists with machine learning expertise to continue to increase and expand in the years to come.

This course covers the basic computational aspects of machine learning, at an undergraduate level.

**Purpose:** To introduce students to fundamental concepts and modern techniques in machine learning, and to prepare students for future work in the area

**Objectives:** To provide students with basic knowledge and understanding of both the theory and practice of machine learning, and to train students on the use and application of machine learning ideas, paradigms and techniques

**Goals:** At the end of the course, students should be able to

- describe, explain and differentiate modern machine learning techniques;
- apply existing models and algorithms;
- *(Optional)* identify potential applications; and
- *(Optional)* select appropriate techniques based on the particular characteristics of the domains and applications under consideration.

**Content**

**Organization:** The course format involves formal lectures and discussions. A list of recommended textbooks follows. Additional reading material will be taken from a variety of sources, including other textbooks in machine learning and related areas, tutorials and research literature in the area, as appropriate for an undergraduate level course.
Some Recommended Textbooks: (Roughly in order of complexity)


Tentative List of Topics

- **Machine-learning fundamentals**: classification, regression and clustering; noisy, noise-free and incomplete data; supervised and unsupervised learning; learning under uncertainty; hypothesis classes, model complexity, model selection, Ockham’s razor and the bias-variance dilemma; generative vs. discriminative probabilistic models; dynamic environments, reinforcement learning and the exploration-exploitation dilemma

- **Basic models and methods**: nearest neighbors, decision trees, linear discrimination, neural networks, support vector machines (SVMs), boosting and bagging, naive Bayes classifiers, gradient-descent, Q-learning

- **Advanced topics**: expectation-maximization (EM), Hidden Markov Models (HMMs), K-means clustering, mixture-of-Gaussians, component analysis

**NOTE:** The list of topics, as well as the emphasis on each topic, may vary slightly depending on the background and interests of the course participants.

Assessment

Students will complete a number of regular homework assignments. Students can complete an optional semester-long course project to receive extra credit for the course.

**Course Project (Optional):** Students complete a project on an application of machine learning to a particular problem, which the students choose in consultation with the instructor. The chosen course project requires instructor’s approval. The project should have an experimental component. Ideally, the project will address a new problem and produce a novel application. Students make an oral presentation of their project proposal. To monitor the project’s development, students periodically make oral presentations as the project progresses. Students produce (and hand in) a final written report on their project and give a final project presentation by the end of the course term.
Student Evaluations: Students’ performance on the homework assignments and their participation in class discussions determine their final grade. The quality of their *optional* project and respective presentations determine the amount of extra credit.

Grades: The following table shows the amount and weighting of each evaluation component in the course.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class participation</td>
<td>20%</td>
</tr>
<tr>
<td>Homework</td>
<td>80%</td>
</tr>
<tr>
<td>Project (Optional)</td>
<td>35%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100% (+ 35% Optional)</strong></td>
</tr>
</tbody>
</table>

Final grades will be assigned using the traditional grading scale (90-100=A, 80-89=B, 70-79=C, 60-69=D, 0-59-F), with deviations at the instructor’s discretion. *Students must submit all the required coursework in order to receive a final grade for the course.*

General University Statements

Americans with Disabilities Act: If you have a physical, psychological, medical or learning disability that may impact your course work, please contact Disability Support Services, ECC (Educational Communications Center) Building, room128, (631) 632-6748. They will determine with you what accommodations, if any, are necessary and appropriate. All information and documentation is confidential.

Academic Integrity: Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person’s work as your own is always wrong. Faculty are required to report any suspected instances of academic dishonesty to the Academic Judiciary. Faculty in the Health Sciences Center (School of Health Technology & Management, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures. For more comprehensive information on academic integrity, including categories of academic dishonesty, please refer to the academic judiciary website at [http://www.stonybrook.edu/uaa/academicjudiciary/](http://www.stonybrook.edu/uaa/academicjudiciary/)

Critical Incident Management: Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of Judicial Affairs any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students’ ability to learn. Faculty in the HSC Schools and the School of Medicine are required to follow their school-specific procedures.