CPSC 4430 Introduction to Machine Learning

CATALOG DESCRIPTION

Course Symbol: CPSC 4430
Title: Machine Learning
Hours of credit: 3

Course Description

Machine learning uses interdisciplinary techniques such as statistics, linear algebra, optimization, and computer science to create automated systems that can sift through large volumes of data at high speed to make predictions or decisions without human intervention. Machine learning as a field is now incredibly pervasive, with applications spanning from business intelligence to homeland security, from analyzing biochemical interactions to structural monitoring of aging bridges, and from emissions to astrophysics, etc. This class will familiarize students with a broad cross-section of models and algorithms for machine learning, and prepare students for research or industry application of machine learning techniques.

Prerequisite:

CPSC 3200, and MATH 2200 and (MATH 2100 or MATH 3100 or MGT 2120 or ENGR 2220)

Overview

Based on fundamental knowledge of computer science principles and skills, probability and statistics theory, and the theory and application of linear algebra. This course provides a broad introduction to machine learning and statistical pattern recognition.

Topics include: (1) supervised learning (generative/discriminative learning, parametric/non-parametric learning, neural networks, and support vector machines); (2) unsupervised learning (clustering, dimensionality reduction, kernel methods); (3) learning theory (bias/variance tradeoffs; VC theory; large margins); and (4) reinforcement learning and adaptive control.

The course will also discuss recent applications of machine learning, such as to robotic control, data mining, autonomous navigation, bioinformatics, speech recognition, and text and web data processing.

TextBook:

Required:
- Ethem Alpaydin, Introduction to Machine Learning, Second Edition , http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=12012. This book will cover all the material in the course.

Recommended:
- Stephen Marsland, Machine Learning: An Algorithmic Perspective. http://www.amazon.com/Machine-Learning-Algorithmic-Perspective-Recognition/dp/1420067184 .
- Christopher M. Bishop, Pattern Recognition and Machine Learning. http://research.microsoft.com/en-us/um/people/cmbishop/prml/.
- Tom Mitchell, Machine Learning, http://www.cs.cmu.edu/~tom/mlbook.html.

Learning Outcomes

By the end of the course, students should be able to:

- Develop an appreciation for what is involved in learning models from data.
- Understand a wide variety of learning algorithms.
- Understand how to evaluate models generated from data.
- Apply the algorithms to a real-world problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models.

Course Schedule

Week 1	• Class overview: Class organization, topics overview, software etc. • Introduction: what is ML; Problems, data, and tools; Visualization; Matlab (I)
Week 2	• Linear regression; SSE; gradient descent; closed form; normal equations; features • Overfitting and complexity; training, validation, test data, and introduction to Matlab (II)
Week 3	• Classification problems; decision boundaries; nearest neighbor methods
Week 4	• Probability and classification, Bayes optimal decisions • Naive Bayes and Gaussian class-conditional distribution
Week 5	• Linear classifiers • Bayes' Rule and Naive Bayes Model
Week 6	• Logistic regression, online gradient descent, Neural Networks • Decision tree and Review for Mid-term
Week 7	• Mid-term
Week 8	• Ensemble methods: Bagging, random forests, boosting • A more detailed discussion on Decision Tree and Boosting
Week 9	• Unsupervised learning: clustering, k-means, hierarchical agglomeration
Week 10	• Advanced discussion on clustering and EM
Week 11	• Latent space methods; PCA.
Week 12	• Text representations; naive Bayes and multinomial models; clustering and latent space models

Week 13	• VC-dimension, structural risk minimization; margin methods and support vector machines (SVM)
Week 14	• Support vector machines and large-margin classifiers • Time series; Markov models; autoregressive models

COURSE REQUIREMENTS:

1. Regular class attendance is the main requirement of this course.

2. Active class participation, this means you must spend some quality time preparing for your next class.

3. Programming assignments, homework, and reports of hands-on labs must be turned in on time when they are due. Unfinished programs and non-working programs turned in on time will be graded; however, assignments not turned in on the due date will **NOT** be accepted. This means that you should start early to work on your programming assignments. Programs must be well-documented to be understood by a novice programmer.

4. Short quizzes may be given without prior notice and **there will be no making up of missed quizzes**.

5. Two examinations and a final examination will be given. There will be **NO** make up for missed exams.

6. You will be issued with one computer account for this class. You have a responsibility and an obligation to prevent abuse and misuse of the university computer resources. Please read the UTC Computer Use Code of Conduct.

7. Individual extra credit assignments for the purpose of propping up a bad grade will **NOT** be given.

Notes taking is encouraged. Notes for the class can be found at the bb5.utc.edu. You can also check the website of the textbook for more resources.

Method of Evaluation

90+ = A; 80-89 = B; 70-79 = C; 60-69 = D; below 60 = F.

Assignments	30%
Course Presentation	10%
Course Project	20%
Exams (1 & 2)	20%

Final	20%
Total	100%

Justification

This course is proposed to meet a growing business need of individuals skilled in artificial intelligence, data analytics, statistical programming and other software skills. The proposed course will combine theory and practice to enable the student to gain the necessary knowledge to compete in the ever changing work environment. Students will learn concepts and methods designed to improve the business decision-making process by putting targeted information into the hands of those who need it most. They will understand business critical processes that drive organizational structures and systems within the context of varying stakeholder interests. Additionally, they will be able to define and evaluate potential initiatives that best fit organizational goals. This proposal will enable UTC to meet the need of local industry such as Blue Cross Blue Shield, U.S. Xpress, UNUM, etc., and educate professionals in areas of business intelligence and data analytics. Specifically, at the end of this course students should be able to effectively develop, manage, integrate, and use corporate information resources. Specifically, on successful completion of this unit, students will be able to:

- source and access data from a variety of databases;
- select and apply appropriate tools for data visualization;
- select and apply descriptive data analytics methods;
- select and apply predictive data analytics methods;
- fit statistical models;
- use the results to produce business intelligence in a variety of settings;
- discover trends in analytical data stores using the data mining techniques of clustering, segmentation, association, and decision trees; and
- present data visually for clear communication to a managerial audience.

Resources

This proposal does not require any additional resources such as staff support, financial resources or physical facilities from the Department of Computer Science and Engineering or any other departments or programs. This course will allow us to more fully utilize the resources we have already in the department. We have the capacity to enroll more graduate students than we presently have, and this concentration will help us recruit more graduate students and utilize our capacity.

Contact

Yu Liang, Yu-Liang@utc.edu

Planned Frequency

Spring semesters

Explanation of Duplication

There is NO duplication or overlapping of proposed course content with courses offering from other departments.

ADA STATEMENT:

Attention: If you are a student with a disability (e.g. physical, learning, psychiatric, vision, hearing, etc.) and think that you might need special assistance or a special accommodation in this class or any other class, call the Office for Students with Disabilities at 425-4006, come by the office - 102 Frist Hall or see http://www.utc.edu/OSD/

If you find that personal problems, career indecision, study and time management difficulties, etc. are adversely affecting your successful progress at UTC, please contact the Counseling and Career Planning Center at 425-4438 or http://www.utc.edu/Administration/CounselingAndCareerPlanning/.