Proposal to Dual-List Com S 574 with Com S 474

1. Full catalog information for each course to be dual-listed, including the course numbers (or proposed course numbers), title, credits, semester offering (if applicable), prerequisites, and description. Dual-listed courses bear common numbers, e.g., 580 (480).

   COM S 574: Introduction to Machine Learning and its Applications
   (Dual-listed with COM S 474) (3-1) Cr. 3. F.
   Prereq: COM S 230 or CPR E 310, STAT 230 or STAT 330
   Introduction to tools and techniques of machine learning for applications. Selected machine learning techniques in practical data mining for classification, regression, and clustering, e.g., association rules, decision trees, linear models, Bayesian learning, support vector machines, artificial neural networks, instance-based learning, Bayesian networks, ensemble learning, and clustering algorithms. Selected applications in data mining, pattern recognition, text and language processing, internet-based information systems, business, and bioinformatics and computational biology.

   COM S 474: Introduction to Machine Learning and its Applications
   (Dual-listed with COM S 574) (3-1) Cr. 3. F.
   Prereq: COM S 230 or CPR E 310, STAT 230 or STAT 330
   Introduction to tools and techniques of machine learning for applications. Selected machine learning techniques in practical data mining for classification, regression, and clustering, e.g., association rules, decision trees, linear models, Bayesian learning, support vector machines, artificial neural networks, instance-based learning, Bayesian networks, ensemble learning, and clustering algorithms. Selected applications in data mining, pattern recognition, text and language processing, internet-based information systems, business, and bioinformatics and computational biology.

2. Graduate faculty status of the proposed instructor.

   Jin Tian, Associate Professor, Department of Computer Science.

3. Number of the dual-listed course credits the department will permit to be used to meet the requirements for an advanced degree. This limit includes dual-listed courses taken in all departments.

   The Department of Computer Science does not have a limit regarding dual listed courses used to meet advanced degree requirements.

4. The differential expectations for graduate students and undergraduates. What additional work will be required for graduate students enrolled in the course? Please describe this work, not in abstract terms (such as "more in-depth participation") but in terms of concrete measurable outcomes or other tangible evidence. Welcome inclusions: specific examples of the additional assignments with details about paper length; the number of additional readings; the length and frequency of oral presentations; portfolio expectations; indications of how these graduate requirements are weighted in the course grade (ex. 40% of final grade); comparisons with undergraduate expectations.

- Graduate students are required to do a term project on a topic approved by the instructor. Graduate students are required to submit a paper and to give an oral presentation on the project at the end of the semester.
- Graduate students have additional questions on problem sets assignments.
- Graduate students have additional questions on the final exam.
- The grades of graduate students will be based on problem sets (25%), laboratory assignments involving programming (25%), a term project requiring written paper and oral presentation (20%), Final exam (25%), and participation in class (5%).
- The grades of undergraduate students will be based on problem sets (35%), laboratory assignments involving programming (30%), Final exam (30%), and participation in class (5%).

5. Reason(s) the course is considered sufficiently rigorous and of such an advanced nature as to challenge graduate students.

   In recent years machine learning has become a very active field with many commercial and scientific applications. The course aims to provide an introduction to the tools and techniques of machine learning used in practical data mining for finding and describing structural patterns in data. This course is targeted to graduate students in computer science and other disciplines (such as bioinformatics and computational biology, human-computer interaction, engineering, statistics, Business and related disciplines) who are interested in applying machine learning techniques to analyze and understand their data.

6. Academic advantages and disadvantages accruing to graduate students taking this course with undergraduates.

   Graduate students taking this course with undergraduates may benefit from a broader diversity of backgrounds and experience.

7. The place of the course in a graduate student’s program of study and why it is not considered a "remedial" undertaking intended to overcome deficiencies in the student’s preparation for graduate work.

   Com S 574 will be a graduate course in the Artificial Intelligence and Machine Learning Breadth Area for Computer Science graduate students. Each Computer Science M.S. student must take at least 3 courses from 2 different Breadth Areas (9 cr.) with a minimum GPA for Core and Breadth Area of 3.0. Each Computer Science Ph.D. student must take at least 4 courses from 2 different Breadth Areas (12 cr.) with a minimum GPA for Core and Breadth Area of 3.0.

   This course is targeted to graduate students in computer science and other who are interested in applying machine learning techniques to analyze and understand their data.

8. The role of the course in an undergraduate’s degree program and the academic qualifications undergraduates must have to take this course.

   Com S 474 is an elective in the Computer Science undergraduate program. Each computer science student must take at least 6 credits of 400-level courses with a grade of C- or better.

The prerequisites of ComS 474 are COM S 230 or CPR E 310, and STAT 230 or STAT 330.

9. The name of the person writing the proposal.

   Jin Tian

ComS 474/574: Introduction to Machine Learning and its Applications
Department of Computer Science
Iowa State University

Instructor

Jin Tian (Email: jtian@iastate.edu; phone: 515-294-8433; Office: 108 Atanasoff Hall)

Course Prerequisites

- Basic knowledge of probability theory and discrete mathematics as applied to computer science. (Necessary mathematics will be reviewed as they arise.)

These topics are covered in COM S 230 or CPR E 310, STAT 230 or STAT 330. In addition, graduate students are expected to have the writing and presentation skills necessary for preparing written reports and presentations based on term projects.

If you are not sure whether you have the necessary background, please talk to the instructor.

Target Audience

This course is targeted to senior undergraduate students and graduate students in computer science and other disciplines (such as bioinformatics and computational biology, human-computer interaction, engineering, statistics, Business and related disciplines) who are interested in applying machine learning techniques to analyze and understand their data.

Course Objectives

In recent years machine learning has become a very active field with many commercial and scientific applications. Machine learning provides the technical basis of data mining to extract useful information hidden in the data. The course aims to provide an introduction to the tools and techniques of machine learning used in practical data mining for finding and describing structural patterns in data. Upon successful completion of the course, students will have an understanding of the different types of tasks that can be addressed through data mining, including association rules, classification, regression, and clustering. Students will learn a variety of machine learning techniques that can be used to solve the different tasks, including association rules, decision trees, linear models, Bayesian classifiers, support vector machines, artificial neural networks, instance-based learning, Bayesian networks, ensemble learning, and clustering algorithms. Students will learn to select appropriate approaches to particular problems and to compare and evaluate the results of different techniques.

The emphasis of the course is on practical data mining using machine learning tools, rather than mathematical theory or advanced details of particular algorithms. Students will work with real world data sets and experiment with a variety of machine learning techniques using an open source data mining tool Weka (http://www.cs.waikato.ac.nz/ml/weka/downloading.html). Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Course Learning Outcomes

- Appreciation of fundamental problems in machine learning
- Ability to make intelligent choices from among available algorithms subject to specific design and performance constraints, and when needed, design variants of existing algorithms.
- Ability to design, implement and evaluate intelligent agents for representative machine learning problems – e.g., learning classification rules from data, etc.
- Familiarity with some current applications of machine learning
- Ability to communicate effectively about machine learning problems, algorithms, implementations, and their experimental evaluation.

Textbook

- Data Mining: Practical Machine Learning Tools and Techniques (Third Edition) by Ian H. Witten, Eibe Frank, and Mark A. Hall, 2011

The course will sometimes draw upon additional references and readings to supplement the treatment of topics available in the primary textbook.

Topics

The following gives a tentative list of topics to be covered in the course (not necessarily in the order in which they will be covered):

- What's data mining and machine learning about
- Getting started with Weka
- Data preparation
- Data mining tasks
- Machine learning techniques

    - Mining association rules
    - Decision trees
    - Linear models
    - Bayesian classifiers
    - Neural networks
    - Support vector machines
    - Instance-based learning
    - Bayesian networks
    - Ensemble learning
    - Unsupervised learning: Clustering
- Evaluating what’s been learned
- Data transformation: attribute selection, discretization, cleansing
- Selected applications in data mining, pattern recognition, text and language processing, internet-based information systems, business, and bioinformatics and computational biology.

Grading Policy for ComS 474

The grades in ComS 474 will be based on problem sets, laboratory assignments (involving programming), final exam, and participation in class. These components will be weighted as follows in assigning an overall numeric score:

- Problem Sets: 35%
- Laboratory assignments: 30%
- Final Exam: 30%
- Participation in class: 5%

Students are guaranteed to receive the letter grade based on the scales shown below. However, the instructor reserves the right to modify the grading scale so as to improve the letter grade if warranted by the circumstances (e.g., unusually high level of difficulty of problem sets).

Com S 474

- 90% - 100% A
- 85% - 89% A-
- 80% - 84% B+
- 75% - 79% B
- 70% - 74% B-
- 65% - 69% C+
- 60% - 64% C
- 55% - 59% C-
- 50% - 54% D+
- 45% - 49% D
- 40% - 44% D-

- <40% F

Grades may be appealed for ONE WEEK after they are distributed (except the final exam which will be by request only). After the appeal period has expired, grade change requests may be denied.

Grading Policy for ComS 574

Additional work will be required for students in ComS 574

- Graduate students are required to do a term project on a topic approved by the instructor. Graduate students are required to submit a paper and to give an oral presentation on the project at the end of the semester.
- Graduate students have additional questions on problem sets assignments.
- Graduate students have additional questions on the final exam.

The grades in ComS 574 will be based on problem sets, laboratory assignments (involving programming), a term project (requiring both written and oral reports), Final exam, and participation in class. These components will be weighted as follows in assigning an overall numeric score:

- Problem Sets: 25%
- Laboratory assignments: 25%
- Term project: 20%
- Final Exam: 25%
- Participation in class: 5%

Students are guaranteed to receive the letter grade based on the scales shown below. However, the instructor reserves the right to modify the grading scale so as to improve the letter grade if warranted by the circumstances (e.g., unusually high level of difficulty of problem sets).

Com S 574

- 93% - 100% A
- 88% - 92% A-
- 83% - 87% B+
- 78% - 82% B
- 73% - 77% B-
- 68% - 72% C+
- 63% - 67% C
- 58% - 62% C-
- 53% - 57% D+
- 48% - 52% D

- 43% - 47% D-
- <43% F

Grades may be appealed for ONE WEEK after they are distributed (except the final exam which will be by request only). After the appeal period has expired, grade change requests may be denied.

Policy on Late Submission of Assignments

Laboratory 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. Programs 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.

Problem Sets

There is a late penalty of 10% of the grade per day up to a maximum of 4 days from the specified due date. Problem sets that are turned in later than 4 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.

The staffs reserve the right to assign the grade on a problem set based on a randomly selected subset of the assigned problems.

Term Projects

Term projects (including code, data, and written report) are to be turned in no later than the due date. Rare exceptions to this policy might be made, at the discretion of the course staff, under demonstrably extenuating circumstances, resulting in the assignment of an 'incomplete' grade.

Class discussion forum

We will be using Piazza for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the instructor, and the TA. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza. If you have any problems or feedback for the developers, email team@piazza.com.

Academic Dishonesty

The class will follow Iowa State University's policy on academic dishonesty (http://www.dso.iastate.edu/ja/academic/misconduct.html). Anyone suspected of academic dishonesty will be reported to the Dean of Students Office. Any student found responsible for academic misconduct will receive a failing grade (F) in the course. The dean of students may impose additional actions (ranging from a disciplinary reprimand to expulsion from the university).

Discussion of general concepts and questions concerning the homework and laboratory assignments among students is encouraged. However, each student is expected to work on the solutions individually (except in the case of assignments that are explicitly assigned to teams of students).

Laboratory Assignments

When discussing code with other students, you may:

- discuss algorithms, data structures, and implementation strategies
- assist in debugging, possibly by suggesting diagnostic print statements or test cases
- provide or receive help in understanding the code that is supplied to the class

It is expected that you have written EVERY LINE OF CODE that you submit (with the exception of code given out in class) as part of your solution for a lab assignment. The following are examples of activities that are PROHIBITED:

- Writing code with another student
- Copying code from another student
- Sharing code with another student (via email, printouts, web, ftp sites, etc.)
- Posting code in a location that is accessible to others
- Using code fragments provided by other students (including students who had taken the course in the past)
- Using code fragments that are freely available (e.g., in public repositories) without properly acknowledging and citing the source

Problem Sets

When discussing problems from assigned problem sets with other students, you may:

- discuss the material presented in class or included in the assigned readings needed for solving the problem(s)
- assist another student in understanding the statement of the problem (e.g., you may assist a non-native speaker by translating some English phrases unfamiliar to that student)

It is expected that you have independently arrived at solutions that you turn in for problem sets. The following are examples of activities that are PROHIBITED:

- sharing solutions or fragments of solutions (via email, whiteboard, handwritten or printed copies, etc.)
- posting solutions or fragments of solutions in a location that is accessible to others
- using solutions or fragments of solutions provided by other students (including students who had taken the course in the past)
- using solutions or solution fragments obtained on the Internet or from solution manuals for text books

Disability Accommodation

Iowa State University complies with the Americans with Disabilities Act and Sect 504 of the Rehabilitation Act. If you have a disability and anticipate needing accommodations in this course, please contact the instructor to set up a meeting within the first two weeks of the semester or as soon as you become aware of your need. Before meeting with the instructor, you will need to obtain a SAAR form with recommendations for accommodations from the Disability Resources Office, located in the Student Services Building. Their telephone number is 515-294-7220 or email disabilityresources@iastate.edu. Retroactive requests for accommodations will not be honored.

Dead Week

This class follows the Iowa State University Dead Week policy as noted in section 10.6.4 of the Faculty Handbook http://www.provost.iastate.edu/resources/faculty-handbook .

Harassment and Discrimination

Iowa State University strives to maintain our campus as a place of work and study for faculty, staff, and students that is free of all forms of prohibited discrimination and harassment based upon race, ethnicity, sex (including sexual assault), pregnancy, color, religion, national origin, physical or mental disability, age, marital status, sexual orientation, gender identity, genetic information, or status as a U.S. veteran. Any student who has concerns about such behavior should contact his/her instructor, Student Assistance at 515-294-1020 or email dso-sas@iastate.edu, or the Office of Equal Opportunity at 515-294-7612.

Religious Accommodation

If an academic or work requirement conflicts with your religious practices and/or observances, you may request reasonable accommodations. Your request must be in writing, and your instructor or supervisor will review the request. You or your instructor may also seek assistance from the Dean of Students Office or the Office of Equal Opportunity.

Contact Information

If you are experiencing, or have experienced, a problem with any of the above issues, email academicissues@iastate.edu.