MACHINE LEARNING FOR DATA INFORMATICS

Time: Wednesdays 2-4:50pm
Classroom: WPH207
Instructors: Professor Stefan Scherer, scherer@ict.usc.edu
Office Hours: ICT 333 or PHE 514/516 (upon request only)

Recommended preparation:

Introduction and Purposes

Learning objectives for students are:

1. Broadly understand major algorithms used in machine learning.
2. Understand supervised and unsupervised learning techniques.
3. Understand Bayesian decision theory and nonparametric methods.
4. Understand Decision trees, dimensionality reduction, clustering, kernel machines.
5. Understand reinforcement learning, Bayesian estimation, hidden Markov models and graphical models.

This is a foundational course with the primary application to data analytics, but is intended to be accessible both to students from technical backgrounds such as computer science, computer engineering, electrical engineering, or mathematics; and to students from less technical backgrounds such business administration, accounting, various medical specializations including preventative medicine and personalized medicine, genomics, and management information systems. A basic understanding engineering and/or technology principles is strongly encouraged, including basic programming skills; as is sufficient mathematical background to provide students with facility in probability, statistics, and linear algebra.

The primary teaching methods will be discussion, case studies, and lectures. Students are expected to perform directed self-learning outside of class which encompasses, among other things, a considerable amount of literature review.

There will be one mid-term exam and a comprehensive final exam. Students will be required to complete six homework assignments, which should average between four to six hours each to complete. The course will be complemented with a mini-project at then end of the term which can be conducted in groups of up to five students. There will be no laboratory assignments, and no special computing facility, hardware or software will be necessary for this course.

Assignments/Reports:
All six homework assignments and project progress are to be submitted individually and students will receive individual scores. However, students may work in groups to complete the tasks. There are one midterm test and a final exam for which date will be posted in the online Schedule of Classes.

Guidelines and additional information will be developed to provide a common vernacular for the assignments. It is crucial that students turn in whatever they have completed on the due date. NO assignment will be accepted late. An incompletes grade will be granted only under the conditions called out in the student handbook, SCAMPUS, which is available online, http://scampus.usc.edu.

Class Communication:
Blackboard at USC will be used for class communication.

Course format

Each class will be three hours including two short pauses. The first two hours will consist of lectures given by Prof. Scherer or one of the guest lecturers. The last hour will be a discussion about the assigned research papers and homework.

Course Material

Required:
- Reading material will be based on published technical papers available via the ACM/IEEE/Springer digital libraries or freely available online. All USC students have automatic access to these digital archives.
- Matlab software, using USC license (for practical exercises)
- All books, papers or reports will be available to students in one of three ways: 1) in the USC bookstore; 2) via USC Blackboard; and/or 3) via the web.
- Ethem Alpaydın, "Introduction to Machine Learning, Second Edition," MIT Press, 2010.

Course Topics and Readings

** Topics and readings are subject to change **

<table>
<tr><th>Classes</th><th>Lectures
(2-3:50pm)</th><th>Readings for discussion sessions and homework
(4-4:50pm)</th></tr>
<tr><td>Week 1</td><td>Introduction to Machine Learning
• Introduction to Machine Learning
• Machine Learning examples
• Introduction to Matlab (1/2)</td><td></td></tr>
<tr><td>Week 2</td><td>Supervised Learning
• Introduction to supervised learning
• PAC learnable
• VC dimension
• Machine Learning Experiments
• Introduction to Matlab (2/2)</td><td>Introduction
• Alpaydin, Ch. 1 and Appendix A
• Assignment 1 due
• NewYorker article – We know how you feel.</td></tr>
<tr><td>Week 3</td><td>Nonparametric Methods
• Bayesian Decision Theory
• Nonparametric methods
• Data “Hands-on”</td><td>Supervised Learning
• Alpaydin, Ch. 2
• Assignment 2 due
• Useful things to know about machine learning - Domingos</td></tr>
</table>

Week 4	**Decision Trees** • Introduction to decision trees • Entropy measures	**Nonparametric Methods** • *Alpaydin, Ch. 3 (3.1 to 3.5) and Ch. 8 (8.1 to 8.5)* • *Assignment 3 due* • *How to display data badly – Howard Wainer*
Week 5	**Linear Discrimination and Neural Networks** • Linear Discrimination • Perceptron • Multilayer Perceptron • Recurrent Neural Networks	**Decision Trees** • *Alpaydin, Ch. 9 and Ch. 19 (19.1, 19.5-19.7)*
Week 6	**User study setup and Evaluation** • Study setup • Statistical Testing	**Linear Discrimination and Neural Networks** • *Alpaydin, Ch. 10 and Ch. 11 (11.1-11.8.2)*
Week 7	**Unsupervised Learning** • Unsupervised Learning • Clustering • Competitive Learning • Semi-supervised Learning	**User study setup and Evaluation** •

Week 8	**Parametric and Multivariate Methods** • Parametric and Multivariate Methods • Regression • Evaluating Regression	**Unsupervised Learning** • *Alpaydin, Ch. 7*
Week 9	**Midterm** • Midterm	**Parametric and Multivariate Methods** • *Alpaydin, Ch. 4 and Ch. 5*
Week 10	**Spring Recess**	

Week 11	**Kernel Machines** • Support Vector Machines • GMM UBM SVM	
Week 12	**Graphical Models** • Bayesian Networks • Belief Propagation	**Kernel Machines** • *Alpaydin, Ch. 13*
Week 13	**Hidden Markov Models** • Hidden Markov Models Principles • Viterbi Algorithm • Applications of HMM	**Graphical Models** • *Alpaydin, Ch. 14*
Week 14	**Deep Learning and Combining Learners** • Auto-encoders • Restricted Boltzmann Machines • Convolutional Neural Networks	**Hidden Markov Models** • *Alpaydin, Ch. 15*
Week 15	**No Lecture - Traveling**	
Week 16	**Review Course**	
Week 17	**Final Exam** **2:4PM please confirm here:** **http://classes.usc.edu/term-20151/finals/** **LVL 17**	
Week 18	**Course project due**	

Grades

- **Grading breakdown**
  - Reading assignments 15%
  - Practical exercises 30%
  - Course project 15%
  - Active participation 10%
  - Exams 30% (10% Midterm; 20% Final Exam)
- **Reading assignments**
  - The reading assignment for each class will consist of 1-2 research papers (posted online at least one week before the class). These papers are specially selected to complement the lectures and show state-of-the-art research.
  - Sunday before each class, 1-3 questions will be posted online.
  - Students must send their answers before the class.
- **Group discussions**
  - Since all students are expected to read the research papers, the discussion should bring something new and interactive to the class. This includes: example datasets, simple implementation of the algorithms, demo, new challenging questions and applications.
- **Practical exercises – Homework Assignments**
  - The practical exercises will be designed to give hands-on experience with machine learning (e.g., SVM, HMM, CRF).
  - Students will need to submit their code (zip files) with their answer to each practical exercise.
- **Course project:**
  - Students can perform the project individually or in team of up to 5.
  - Teams will participate in one or more of the available challenges on https://www.kaggle.com/competitions

Statement for Students with Disabilities

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to TA) as early in the semester as possible. DSP is located in STU 301 and is open 8:30 a.m.–5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776.

Statement on Academic Integrity

USC seeks to maintain an optimal learning environment. General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that individual work will be submitted unless otherwise allowed by an instructor, and the obligations both to protect one's own academic work from misuse by others as well as to avoid using another's work as one's own. All students are expected to understand and abide by these principles. *Scampus,* the Student Guidebook, contains the Student Conduct Code in Section 11.00, while the recommended sanctions are located in Appendix A: http://www.usc.edu/dept/publications/SCAMPUS/gov/. Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty. The Review process can be found at: http://www.usc.edu/student-affairs/SJACS/.