USC Viterbi
School of Engineering

INF 552 Machine Learning for Data Informatics

Units: 4

Spring 2017—Monday & Wednesday—Time: 10-11:50 AM

Location: VKC 260

Instructor: Stefan Scherer

Office: ICT 338

Office Hours: by appointment only

Contact Info: scherer@ict.usc.edu

Teaching Assistant: Lixing Liu

Office:

Office Hours:

Contact Info: lixingli@usc.edu

IT Help: TBA

Hours of Service:

Contact Info:

Course Description
Practical applications of machine learning techniques to real-world problems. Uses in data mining and recommendation systems and for building adaptive user interfaces.

Expanded Description
Machine learning techniques allow computers to act without being explicitly programmed. These techniques learn from examples or experience rather than from explicit rules. Machine learning has practical value in many application areas of computer science, such as data mining, recommendation systems, and building adaptive user interfaces. This class will focus on practical and effective applications of a wide range of machine learning techniques to a variety of real-world problems.

This class will be primarily individual study, with weekly assigned readings, eight homework assignments, one midterm exam, and a final.

Learning Objectives
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.

Course Notes

Methods of Teaching: 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 eight homework assignments, which should average between four to six hours each to complete. There will be no laboratory assignments, and no special computing facility, hardware or software will be necessary for this course.

Assignments/Reports: All eight homework assignments are to be submitted individually and students will receive individual scores. However, students may work in groups to complete the tasks. There are a midterm 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.

Required Readings and Supplementary Materials

All books, papers or reports will be available to students in one of three ways: 1) in the USC bookstore; 2) via a CD that the instructor will provide at the beginning of class; and/or 3) via the web.

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.
- Python, 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.

Grades

- **Grading breakdown**
  - Reading assignments 20%
  - Practical exercises 25%
  - Course project 25%
  - 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.
  - 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

Course project: the purpose of the class project is for you to learn hands-on experience of identification of a data problem and to apply machine learning approaches in order to solve it. Students are encouraged to identify unique applications for machine learning and develop novel approaches; however sample topics will be provided to students. Working as a group is permitted if the project is large enough to justify this. A team can consist of up to 3 persons.

Project Timeline (changes might apply):

- Week 1 - 4: Identifying team members and project topics
- Week 5: Draft proposal due (team member, topics and milestones)
- Week 7: Proposal due

- Week 11: Mid-term report due (data description, preliminary results)
- Week 15: Project presentation
- Week 16: Final report due (task and model description, major discovery, lessons learned)

Sample projects "*Public Speaking Performance Assessment of Political Debaters:*" the goal of the project is to develop an automatic multimodal model of the performance of public debaters. Students can easily find resources available online, e.g. Presidential debates on Youtube and machine learning toolboxes as well as feature extraction and signal processing repositories. A project of this size usually consists of 2-3 persons. The team will work together on collecting the data, annotating the data, examining the preliminary results, identifying challenges and developing a machine learning solution to the topic.

Grading breakdown of the course project:

- Proposal 10%
- Mid-term report 25%
- Final report 35%
- Presentation 30%

Course Schedule: A Weekly Breakdown

Classes	Lectures	Readings for discussion sessions and homework (Reading list is not complete and is subject to change)
Week 1	**Introduction to Machine Learning** • Introduction to Machine Learning • Machine Learning examples • Introduction to Machine Learning with Python and Matlab (1/2)	
Week 2	**Supervised Learning** • Introduction to supervised learning • PAC learnable • VC dimension • Machine Learning Experiments • Introduction to Machine Learning with Python and Matlab (2/2)	**Introduction** • *Alpaydin, Ch. 1 and Appendix A* • *Assignment 1 due* • *The end of code – Jason Tanz (Wired Mag.)* • *Andrew Ng shares the astonishing ways deep learning is changing the world - https://www.import.io/post/andrew-ng-shares-the-astonishing-ways-deep-learning-is-changing-the-world/* •
Week 3	**Nonparametric Methods** • Bayesian Decision Theory • Nonparametric methods • Data "Hands-on"	**Supervised Learning** • *Alpaydin, Ch. 2* • *Assignment 2 due* • *Domingos, P. (2012). A few useful things to know about machine learning. Communications of the ACM, 55(10), 78-87.*
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)* • *Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9(2579-2605), 85.* • *Wainer, H. (1984). How to display data badly. The American Statistician, 38(2), 137-147.*
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)* • *Assignment 3 due* • *Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1(1), 81-106.* • *Scott, C., & Nowak, R. (2004). On the adaptive properties of decision trees. In Advances in Neural Information Processing Systems (pp. 1225-1232).*
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)* • *Introduction to neural networks (chapter 5) – David Kriesel* • *Hammer, B., & Villmann, T. (2003, April). Mathematical Aspects of Neural Networks. InESANN (pp. 59-72).*
Week 7	**Unsupervised Learning** • Unsupervised Learning • Clustering • Competitive Learning • Semi-supervised Learning	**User study setup and Evaluation** • *Assignment 4 due* • *Risko, E. F., & Kingstone, A. (2011). Eyes wide shut: implied social presence, eye tracking and attention. Attention, Perception, & Psychophysics, 73(2), 291-296.*
Week 8	**Parametric and Multivariate Methods** • Parametric and Multivariate Methods • Regression • Evaluating Regression	**Unsupervised Learning** • *Alpaydin, Ch. 7* • *Qian, F., Pathak, A., Hu, Y. C., Mao, Z. M., & Xie, Y. (2010, June). A case for unsupervised-learning-based*

		spam filtering. In ACM SIGMETRICS Performance Evaluation Review (Vol. 38, No. 1, pp. 367-368). ACM. • *Ji, M., Yang, T., Lin, B., Jin, R., & Han, J. (2012). A simple algorithm for semi-supervised learning with improved generalization error bound. arXiv preprint arXiv:1206.6412.* • *Assignment 5 due*
Week 9	**Midterm** • Midterm	**Parametric and Multivariate Methods** • *Alpaydin, Ch. 4 and Ch. 5*
Week 10	**Kernel Machines** • Support Vector Machines • GMM UBM SVM	
Week 11	**Graphical Models** • Bayesian Networks • Belief Propagation	**Kernel Machines** • *Alpaydin, Ch. 13* • *Scherer, S., Kane, J., Gobl, C., & Schwenker, F. (2013). Investigating fuzzy-input fuzzy-output support vector machines for robust voice quality classification. Computer Speech & Language, 27(1), 263-287.* • *Bennett, K. P., & Campbell, C. (2000). Support vector machines: hype or hallelujah?. ACM SIGKDD Explorations Newsletter, 2(2), 1-13.*
Week 12	**Hidden Markov Models** • Hidden Markov Models Principles • Viterbi Algorithm • Applications of HMM	**Graphical Models** • *Alpaydin, Ch. 14* • *Airoldi, E. M. (2007). Getting started in probabilistic graphical models. PLoS Comput Biol, 3(12), e252.* • *Bishop, C. M. (2013). Model-based machine learning. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 371(1984), 20120222.* • *Assignment 6 due*
Week 13	**Deep Learning and Combining Learners** • Auto-encoders • Restricted Boltzmann Machines • Ensemble Classifiers/Classifier fusion	**Hidden Markov Models** • *Alpaydin, Ch. 15* • *Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257-286.* • *Anderson, B., & Moore, A. (2005, August). Active learning for hidden markov models: Objective functions and algorithms. In Proceedings of the 22nd international conference on Machine learning (pp. 9-16). ACM.* • *Assignment 7 due*
Week 14	**Deep Learning cont'd** • Long Short-term Memory Networks • Convolutional Neural Networks	**Deep Learning and Combining Learners** • *Le, Q. V. (2013, May). Building high-level features using large scale unsupervised learning. In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on (pp. 8595-8598). IEEE.* • *Chatterjee, M., Park, S., Morency, L. P., & Scherer, S. (2015, November). Combining Two Perspectives on Classifying Multimodal Data for Recognizing Speaker Traits. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (pp. 7-14). ACM* • *.Assignment 8 due*

Week 15	**Project Presentation/Course Review**	**Deep Learning and Combining Learners** • *Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).* • *Trigeorgis, G., Ringeval, F., Brueckner, R., Marchi, E., Nicolaou, M. A., Schuller, B., & Zafeiriou, S. ADIEU FEATURES? END-TO-END SPEECH EMOTION RECOGNITION USING A DEEP CONVOLUTIONAL RECURRENT NETWORK.*
	Final Exam	Date: For the date and time of the final for this class, consult the USC *Schedule of Classes* at www.usc.edu/soc.

Statement on Academic Conduct and Support Systems

Academic Conduct

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Section 11, *Behavior Violating University Standards* https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* http://equity.usc.edu or to the *Department of Public Safety* http://adminopsnet.usc.edu/department/department-public-safety. This is important for the safety of the whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. *The Center for Women and Men* http://www.usc.edu/student-affairs/cwm/ provides 24/7 confidential support, and the sexual assault resource center webpage http://sarc.usc.edu describes reporting options and other resources.

Support Systems

A number of USC's schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the *American Language Institute* http://dornsife.usc.edu/ali, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information* *http://emergency.usc.edu* will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.