



INF 552 Machine Learning for Data Informatics
Fall 2019 (4 Units)

Lectures Time: Monday 3:30PM --- 6:50 PM

Location: THH 116

Instructor: Prof. Wei-Min Shen (www.isi.edu/~shen)

Office Hours: After Monday Lectures (by appointment only)

Contact Info: shen@isi.edu (include "INF 552" in the subject)

Teaching Assistant: TBA

Office: TBA

Office Hours: TBA

Contact Info: tbd@usc.edu

IT Help: TBA

Hours of Service:

Contact Info:

Webpages: [USC Blackboard Class Page](#) and [Piazza Class Page](#)

- All HWs, handouts, solutions will be posted in PDF format
- *Student is responsible to stay current with the webpage*

Course Description

Basic understanding of machine learning techniques and practical applications to selected real-world problems in data analysis and data mining.

Machine learning techniques allow computers to act without being explicitly programmed. These techniques can learn from examples or experience rather than acting on explicitly given programs. Machine learning has practical value in many application areas of computer science, such as data mining, autonomous systems, and building adaptive intelligent systems. This class will focus on the understanding of widely used machine learning techniques and their practical and effective applications to a variety of real-world problems.

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, and 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 as 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. Sufficient mathematical background will be needed to provide students with facility in probability, statistics, and linear algebra.

Course Contents

The course will be run as a lecture class with student participation strongly encouraged. There are weekly lectures and reading reports, five programming assignments, one competition, one midterm and one final exam. Students are encouraged to do the readings prior to the discussion in class. All of the course materials, including the reading assignments, lecture notes, and homework assignments, will be posted online.

Technological Proficiency and Hardware/Software Required

Students are expected to know how to program in a language such as Python. Students are also expected to have their own laptop or desktop computer where they can install and run software to do the homework and project assignments.

Text Books and Readings

1. Book (AL): Ethem Alpaydın, *Introduction to Machine Learning*, 3rd Edition, MIT Press, 2014.
2. Selected Book Chapters (WS): Wei-Min Shen, *Autonomous Learning from the Environments*, Computer Science Press, 1994. Download from <http://www.isi.edu/~shen/book.html>
3. Other reading material may be selected from published technical papers available online or via ACM/IEEE/Springer digital libraries accessible for all USC students.

Letter Grade Distribution

≥93.00	A	73.00 - 76.99	C
90.00 - 92.99	A-	70.00 - 72.99	C-
87.00 - 89.99	B+	67.00 - 69.99	D+
83.00 - 86.99	B	63.00 - 66.99	D
80.00 - 82.99	B-	60.00 - 62.99	D-
77.00 - 79.99	C+	≤59.99	F

Grading Schema

Reading reports	8%
Homework programming assignments	50% (10% each)
Midterm Exam	14%
Final Exam	20%
Machine Learning Competition Project	8%
Total	100%

Reading Reports

The reading reports for each week will consist of reading book chapters and/or research papers. Your reports will be 1-2 pages long, with your typed in answers to the following six questions. The reports are due before the next class. Students may skip any two reports you choose during the semester.

1. What is the problem to be solved in this reading?
2. Why is the problem interesting and useful?
3. Why is the problem difficult or challenging to solve?
4. What is the technique presented in the reading (use your own words)?
5. What are the examples used in the reading and how well they work for the technique?
6. Which parts in the reading that you think are difficult to understand?

Homework Programming Assignments (PA)

There will be five (5) programming assignments. The assignments must be done individually and submitted before due time. Each assignment will be graded and the specific rubric for each assignment is given in the assignment. Each submission will be checked for plagiarism. Some example topics for these theoretical and programming exercises are as follows: Decision Trees, Neural Networks, Bayesian Models, Kernel Machines, Clustering, Markov models, and Reinforcement Learning.

Machine Learning Competition Project: There will be a final project based on the topics introduced in class. The final project is to build an advanced machine learning agent and compete with other students and TA for achieving the best performance. The winners will be announced and honored with a good recommendation for future references. There would be some bonus for the top few students for the competition. To enter the final competition, your system must pass some baseline tests.

Exams

- Midterm: October 14, 3:30-5:30PM (in class)
- Final: December 16, 2-4PM (set by the university)

Late Submission Policy:

For any assignment, students will have one chance to be late within a half day, after that there will be -30% to -10% penalty for the grade for each day that is late.

Tentative Schedule

Week	Lectures	Readings (due before next class)	Deliverables/Due Dates
1. (8/26)	Introduction to machine learning	AL Ch. 1 , WS Ch. 1-2	
2. (9/2)	Labor Day (no class)	AL Appendix, WS Ch. 3	
3. (9/9)	Supervised Learning	AL Ch. 2, WS Ch. 4.1-4.5	PA1 assigned
4. (9/16)	Decision Trees	AL Ch. 9, 19.1, 19.5-19.7	
5. (9/23)	Neural Networks (Linear-Discrimination)	AL Ch. 10, 11.1-11.8.2, WS Ch. 4.11	PA1 due (3:00PM), PA2 assigned
6. (9/30)	Bayesian & non-parametric Methods	AL Ch. 3.1-3.5, 8.1-8.5, WS Ch. 4.12	
7. (10/7)	Kernel Machines	AL Ch. 13	PA2 due (3:00PM), PA3 assigned
8. (10/14)	Midterm (in class)		
9. (10/21)	Unsupervised Learning, Clustering	AL Ch. 7, WS Ch. 5.1, 5.8	PA3 due (3:00PM), PA4 assigned
10. (10/28)	Parametric and Multivariable Methods	AL Ch. 4, 5	Competition project assigned
11. (11/4)	Graphic Models	AL Ch. 14	PA4 due (3:00PM), PA5 assigned
12. (11/11)	Hidden Markov Models	AL Ch. 15, WS Ch. 5.10	
13. (11/18)	Reinforcement Learning	AL Ch. 18, WS Ch. 6.11,6.13	PA5 due (3:00PM)
14. (11/25)	Bayesian Estimation	AL Ch. 16	
15. (12/2)	Deep Learning and Combined Learners	AL Ch. 17	
16. (12/9)	*Friday, 12/6 is the last day of class		Competition project due (3:00PM)
Final (12/16)	2-4PM (set by the university)		

Grading Policies

- Students are encouraged to discuss homework and projects with one another, but each student must do their own work and submit individual written/coded solutions in their own hand. Copying the solutions or submitting identical code/answers are written evidence of cheating. The penalty ranges from F on the homework or exam, to an F in the course, to recommended expulsion.
- Posting the homework/project assignments and their solutions to online forums or sharing them with other students is strictly prohibited and infringes the copyright of the instructor. Instances will be reported to USC officials as academic dishonesty for disciplinary action.

Exam Policy

- No make-up exams will be given. If you cannot make the scheduled dates due to conflicts with other classes or personal matter, you must drop the class. In the case of a required business trip or a medical emergency, a signed letter from your manager or physician has to be submitted. This letter must include the contact of your physician or manager.
- Midterm and final exams will be closed book and notes. No calculators are allowed nor are computers and cell-phones or any devices that have Internet capability.
- All exams are cumulative, with an emphasis on material presented since the last exam.

Attendance

- Students are required to attend all the lectures and discussion sessions and actively participate in class. Use of cellphones and laptops for non-class activities is prohibited in the classroom.
- Handouts and study guides will be distributed during classes.

Important Notes

- Textbooks are secondary to the lecture notes and homework/project assignments.

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/academicssupport/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.