DSCI 552 Machine Learning for Data Science

Units: 4
Instructor: Ke-Thia Yao
Contact Info: kyao@isi.edu

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 data science application areas, including biology, physics and finance. This class will provide students with a solid understanding of major machine learning algorithms and practical experience in building and training machine learning models through lectures, homework assignments, program assignments, exams and a semester project.

Learning Objectives

After successful completion of this course students will be able to
- Understand the principles of machine learning
- Understand major algorithms used in machine learning
- Design, implement and train machine learning models on data science datasets

Recommended Preparation

The recommended preparation is for students to have taken courses and/or to have prior knowledge of probability/statistics, linear algebra, some calculus and Python programming language.

Office Hours

Tuesday and Thursday 2:00 PM - 4:00 PM at PHE 514
Also, by appointment: kyao@isi.edu

Readings

The main textbook for this course is:

Alpaydin, Ethem. 2020. *Introduction to Machine Learning*. 4th ed. Adaptive Computation and Machine Learning. Cambridge: MIT Press.

Optional textbook describing the Scikit Learn machine learning toolkit is:

Géron, Aurélien. 2020. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition*. Online version available through USC Library
https://libraries.usc.edu/databases/safari-books

Aslo, the Scikit-Learn website provides excellent documentation and user guides:

https://scikit-learn.org/stable/index.html

In addition reading assignments will be selected from various periodicals and other sources.

Grading Breakdown

Assessment	% of Grade
Homework / Programming Assignments	30%
Midterm 1	25%
Midterm 2	25%
Semester Project	20%

Assignment Submission Policy

Assignments and semester project will be submitted electronically via Blackboard. Assignments will be accepted after the deadline with the following grade penalties:

- 1 day late: lose 10%

- 2 days late: lose 30% (10% + 20%)
- 3 days late: lose 60% (30% + 30%)
- Greater than 4 days late not accepted

No personal emergencies will be entertained (with the exception of the USC granted emergencies, in which case official documents need to be shown).

Participation

Students are expected to actively participate in this course. Participation includes:

- Careful reading and viewing of assigned materials by the date due
- Regular, substantive contributions to discussions and in-class questions
- Active engagement with online content

Course grades for students who do not contribute to the course through active participation in class may be affected.

Programming Assignments

For this course we will be using the popular machine learning toolkit [SciKit Learn](#). The toolkit provides a rich variety of machine learning algorithm implementations. The toolkit is well documented and it provides [user guides](#).

We will use [Jupyter notebooks](#) for the programming assignments.

Course Schedule

Week 1: Overview of Machine Learning and Supervised Learning
Topics:
- What is machine learning
- Machine learning tasks and examples applications
- Machine learning process
- Learning bias
- Linear least squares and k-nearest neighbors classification
Readings:
- Alpaydin Chapter 1
- Alpaydin Chapter 2.1
- Géron Chapter 1
- Géron Chapter 2

Week 2: Supervised learning (cont) and Bayesian Decision Theory
Topics:
- Classification and regression

- VC dimension and PAC learning
- Model selection and generalization
- Train, validation and test set
- Review of probabilities
- Bayes's rule
- Discriminant functions and decision regions
- Association rules

Readings:
- Alpaydin Chapter 2
- Alpaydin Chapter 3
- Alpaydin Appendix A
- Géron Chapter 3

Week 3: Parametric Methods

Topics:
- Maximum likelihood estimation
- Bayes' estimator
- Linear regression, polynomial regression
- Bias/variance dilemma
- Naive Bayes classifier

Readings:
- Alpaydin Chapter 4
- Alpaydin Appendix B
- Domingos, Pedro. 2012. "A Few Useful Things to Know About Machine Learning."
  Comm. ACM 55 (10): 78–87. https://doi.org/10.1145/2347736.2347755.
- Géron Chapter 4.1

Week 4: Nonparametric Methods and Decision Trees

Topics:
- Nonparametric density estimation (histogram, kernel and k-nearest neighbor estimators)
- Smoothing models
- Univariate trees
- Pruning
- Learning rules

Readings:
- Alpaydin Chapter 8
- Alpaydin Chapter 9
- Géron Chapter 6

Week 5: Dimensionality Reduction

Topics:
- Curse of dimensionality
- Feature subset selection
- Principal component analysis and multidimensional scaling

- Singular value decomposition and matrix factorization
- Isomap, locally linear embedding

Readings:
- Alpaydin Chapter 6
- Géron Chapter 8

Week 6: Midterm & Clustering

Topics:
- k-means clustering
- Expectation-maximization algorithm
- Spectral clustering
- Hierarchical clustering

Readings:
- Alpaydin Chapter 7
- Géron Chapter 9

Week 7: Linear Discrimination

Topics:
- Linear discrimination and generalizing the linear model
- Gradient descent
- Logistic regression
- Perceptron and training the perceptron
- XOR problem
- Multilayer perceptrons and the backpropagation algorithm
- Autoencoders and word2vec

Readings:
- Alpaydin Chapter 10
- Géron Chapter 4

Week 8: Multilayer Perceptrons

Topics:
- Multilayer perceptrons and the backpropagation algorithm
- Autoencoders and word2vec

Readings:
- Alpaydin Chapter 11
- Géron Chapter 10

Week 9: Deep Learning

Topics:
- Training multiple hidden layers
- Improving training convergence
- Regularization
- Convolutional neural networks
- Learning sequences

- Generative adversarial network

Readings:
- Alpaydin Chapter 12
- Géron Chapter 11

Week 10: Spring Recess

Week 11: Kernel Machines
Topics:
- Optimal separating hyperplane and soft margin hyperplane
- Support vector machines
- Kernel trick
- Vectorial kernels
- Kernel machines for regression

Readings:
- Alpaydin Chapter 14
- Géron Chapter 5

Week 12: Midterm & Design and Analysis of Machine Learning Experiments
Topics:
- Cross-validation
- Resampling methods
- Measuring and assessing classifier performance

Readings:
- Alpaydin Chapter 20
- Géron Chapter 3

Week 13: Graphical Models and Boltzmann Machines
Topics:
- Conditional Independence and d-Separation
- Belief Propagation
- Markov Random Fields, Boltzmann Machines (BM)
- Restricted BM
- Limited BM and DWave quantum adiabatic annealer

Readings:
- Alpaydin Chapter 15
- https://docs.dwavesys.com/docs/latest/c_gs_1.html
- Liu, Jeremy, Ke-Thia Yao, and Federico Spedalieri. 2020. "Dynamic Topology Reconfiguration of Boltzmann Machines on Quantum Annealers." *Entropy* 22 (11): 1202. https://doi.org/10.3390/e22111202.

Week 14: Hidden Markov Models
Topics:
- Discrete Markov Process

- Hidden Markov Models (HMMs)
- Three basic problems of HMMs
- Viterbi Algorithm

Readings:
- Alpaydin Chapter 15
- Viterbi, A.J. 2006. "A Personal History of the Viterbi Algorithm." *IEEE Signal Processing Magazine* 23 (4): 120–42. https://doi.org/10.1109/MSP.2006.1657823.

Week 15: Combining Multiple Learners

Topics:
- Generating diverse learners
- Voting, Bagging, Boosting
- Random forest and gradient boosting

Readings:
- Alpaydin Chapter 18

Week 16: Project Presentation

Students are expected to present their semester projects to the class, as one component of the semester project, in addition to the one-page project proposal, and the project status report.

Statement on Academic Conduct and Support Systems

Academic Integrity:

The University of Southern California is a learning community committed to developing successful scholars and researchers dedicated to the pursuit of knowledge and the dissemination of ideas. Academic misconduct, which includes any act of dishonesty in the production or submission of academic work, comprises the integrity of the person who commits the act and can impugn the perceived integrity of the entire university community. It stands in opposition to the university's mission to research, educate, and contribute productively to our community and the world.

All students are expected to submit assignments that represent their own original work, and that have been prepared specifically for the course or section for which they have been submitted. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s).

Other violations of academic integrity include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), collusion, knowingly assisting others in acts

of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university. All incidences of academic misconduct will be reported to the Office of Academic Integrity and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see [the student handbook](#) or the [Office of Academic Integrity's website](#), and university policies on [Research and Scholarship Misconduct](#).

Please ask your instructor if you are unsure what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

Students and Disability Accommodations:

USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at [osas.usc.edu](#). You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Support Systems:

[*Counseling and Mental Health*](#) *- (213) 740-9355 – 24/7 on call*

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

[*988 Suicide and Crisis Lifeline*](#) *- 988 for both calls and text messages – 24/7 on call*

The 988 Suicide and Crisis Lifeline (formerly known as the National Suicide Prevention Lifeline) provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week, across the United States. The Lifeline is comprised of a national network of over 200 local crisis centers, combining custom local care and resources with national standards and best practices. The new, shorter phone number makes it easier for people to remember and access mental health crisis services (though the previous 1 (800) 273-8255 number will continue to function indefinitely) and represents a continued commitment to those in crisis.

[Relationship and Sexual Violence Prevention Services (RSVP)](#) - (213) 740-9355(WELL) – 24/7 on call

Free and confidential therapy services, workshops, and training for situations related to gender- and power-based harm (including sexual assault, intimate partner violence, and stalking).

[Office for Equity, Equal Opportunity, and Title IX (EEO-TIX)](#) - (213) 740-5086

Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

[Reporting Incidents of Bias or Harassment](#) - (213) 740-5086 or (213) 821-8298

Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

[The Office of Student Accessibility Services (OSAS)](#) - (213) 740-0776

OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

[USC Campus Support and Intervention](#) - (213) 740-0411

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

[Diversity, Equity and Inclusion](#) - (213) 740-2101

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

[USC Emergency](#) - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

[USC Department of Public Safety](#) - UPC: (213) 740-6000, HSC: (323) 442-1200 – 24/7 on call

Non-emergency assistance or information.

[*Office of the Ombuds*](#) *- (213) 821-9556 (UPC) / (323-442-0382 (HSC)*

A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

[*Occupational Therapy Faculty Practice*](#) *- (323) 442-2850 or* otfp@med.usc.edu

Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.