

# DSCI 522 Machine Learning for Data Science Syllabus

## 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, and Python programming language.

## 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	35%
Class participation	5%
Midterm	20%
Final Exam	20%
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. Cumulative of 10% times number of days late:

- 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

### **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

### **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>.

#### **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

#### **Week 5: Dimension 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

#### **Week 6: Clustering**

Topics:

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

Readings:

- Alpaydin Chapter 7

#### **Week 7: Linear Discrimination and Multilayer Perceptrons**

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
- Alpaydin Chapter 11

### **Week 8: Midterm**

### **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

### **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

### **Week 12: 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](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 13: 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 14: Combining Multiple Learners**

Topics:

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

Readings:

- Alpaydin Chapter 18

### **Week 15: 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 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 Part B, Section 11, "Behavior Violating University Standards" [policy.usc.edu/scampus-part-b](https://policy.usc.edu/scampus-part-b). Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, [policy.usc.edu/scientific-misconduct](https://policy.usc.edu/scientific-misconduct).

## **Support Systems:**

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

[studenthealth.usc.edu/counseling](http://studenthealth.usc.edu/counseling)

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

*National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call*

[suicidepreventionlifeline.org](http://suicidepreventionlifeline.org)

Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

*Relationship and Sexual Violence Prevention and Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7 on call*

[studenthealth.usc.edu/sexual-assault](http://studenthealth.usc.edu/sexual-assault)

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

*Office of Equity and Diversity (OED)- (213) 740-5086 | Title IX – (213) 821-8298*

[equity.usc.edu](http://equity.usc.edu), [titleix.usc.edu](http://titleix.usc.edu)

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. The university prohibits discrimination or harassment based on the following *protected characteristics*: race, color, national origin, ancestry, religion, sex, gender, gender identity, gender expression, sexual orientation, age, physical disability, medical condition, mental disability, marital status, pregnancy, veteran status, genetic information, and any other characteristic which may be specified in applicable laws and governmental regulations. The university also prohibits sexual assault, non-consensual sexual contact, sexual misconduct, intimate partner violence, stalking, malicious dissuasion, retaliation, and violation of interim measures.

*Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298*

[usc-advocate.symplicity.com/care\\_report](http://usc-advocate.symplicity.com/care_report)

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

*The Office of Disability Services and Programs - (213) 740-0776*

[dsp.usc.edu](http://dsp.usc.edu)

Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

*USC Support and Advocacy - (213) 821-4710*

[uscса.usc.edu](http://uscса.usc.edu)

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

*Diversity at USC - (213) 740-2101*

[diversity.usc.edu](http://diversity.usc.edu)

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*

[dps.usc.edu](http://dps.usc.edu), [emergency.usc.edu](http://emergency.usc.edu)

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) 442-120 – 24/7 on call*

[dps.usc.edu](http://dps.usc.edu)

Non-emergency assistance or information.