EE 660 **Course Syllabus** Fall 2020

V1 SOC, 8/24/2020

Machine Learning from Signals: Foundations and Methods

Administrative information

Times and days

Lecture: Tu Th 2:00 – 3:20 PM, online and through DEN@Viterbi.

Discussion session: Friday 3:30 – 4:20 PM, online and through DEN@Viterbi.

Lectures and discussion sessions can be viewed by live streaming, and can be viewed by archived video any time after the live event. The live streaming provides for you optionally to interact via audio, video, text, and other software-provided actions (such as raising one's hand, voting, responding to short-answer questions, etc.).

Catalogue description

Supervised, semi-supervised, and unsupervised machine learning; classification and regression. Model complexity, assessment, and selection; performance (error) on unseen data.

Extended course description

Foundations of machine learning, which apply to many or all algorithmic approaches, will be studied. These will include feasibility of learning; complexity of learning; regularization, overfitting and underfitting of models to data; model selection and assessment; and prediction of performance on unseen data. Particular methods that are key to machine learning from signals will also be covered. These will include linear and nonlinear techniques for regression as well as for classification, in the supervised learning realm. Also, methods described for classification by semi-supervised learning (using some labeled data and some unlabeled data for training), and for clustering by unsupervised learning (using only unlabeled data), will include statistical and distribution-free approaches. Feature selection, including the use of sparsity, will also be studied briefly. Students will be exposed to examples of techniques run on both synthetic and real-word data, through examples in lectures and the reading, as well as in homework problems and in the course project(s).

Learning Objectives

(1) To provide the student with a solid foundation in machine learning principles and the capability to apply them to problems.

(2) To give the student knowledge of common and successful methods (techniques and algorithms) in machine learning, and the ability to use them.

(3) To provide the student with sufficient foundation and knowledge so that he or she can learn about many of the plethora of machine learning techniques that now exist, on his or her own as needed.

Preparation

Prerequisites: EE 503, EE 510/441, and EE 559.

Computer Hardware/Software Requirements:

Students are required to use their choice of either Python (recommended) or MATLAB (also allowed) for all homework computer problems unless stated otherwise in the problem. Some tips and resources for students that want to learn Python will be provided. For the class project(s), students may use MATLAB, Python, and/or C/C+. (To use other languages for the class project, check first with the TA or instructor.)

All students that use Python will be responsible for installing and maintaining their own Python distribution. A recommended source (distribution) is Anaconda:

https://www.anaconda.com/products/individual

MATLAB is available on campus computers; or can be downloaded from:

https://viterbigrad.usc.edu/technical-support/matlab/

Note that for some computer problems, you may be required to code up the problem yourself, without the use of libraries or software packages. Each homework problem will specify what packages are allowed. And for portions of your course project you may find it advantageous to do the coding yourself; for other portions, you may find using functions or methods from a packaged library is preferable.

Textbooks, reading materials, and other resources

Required textbooks and reading materials

Selected portions of each book will be used for the class. Please note that the total cost of the two books is approximately the same as the cost of one typical textbook in a graduate-level EE class.

1. Kevin P. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press, Cambridge, 2012). [In short, "Murphy"] (Available at USC bookstore and online sellers)
2. Yasir S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin, *Learning From Data* (AMLbook.com, 2012). [In short, "AML"] (Available from Amazon)

In addition, other materials will be used in portions of the course, including:

3. Xiaojin Zhu and Andrew B. Goldberg, *Introduction to Semi-Supervised Learning* (Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan and Claypool Publishers, 2009). [In short, "Zhu"] (Available for download through USC Library).
4. Rui Xu and Donald Wunsch II, "Survey of Clustering Algorithms", *IEEE Trans. Neural Networks*, Vol. 16, No. 3 (May 2005). [In short, "Xu"]. A link will be provided on the course web site.

Additional resource books for your information (not required)

i. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition (Springer, 2009)

ii. R. O. Duda, P. E. Hart, and D. G. Stork, , *Pattern Classification,* Second Edition (Wiley-Interscience, John Wiley and Sons, Inc., New York, 2001)

iii. C. M. Bishop, "Pattern Recognition and Machine Learning" (Springer, 2006)

iv. M Mohri, A. Rostamizadeh, and A. Talwalker, *Foundations of Machine Learning*, second edition (MIT Press, Cambridge, 2018)

v. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning* (MIT Press, Cambridge, 2016).

Course web site (Desire2Learn system)

courses.uscden.net

The site includes:

– Links to online lectures, discussion sessions, and office hours.
– Course materials (handouts, homework assignments, lecture notes, lecture videos, etc.), which will be posted as we progress through the semester.
– Link to our discussion forum (piazza).
– Course calendar, showing events and deadlines.
– Grade book, showing your scores on assignments to date.
– Dropboxes for uploading your completed assignments, and links for viewing and retrieving your graded assignments.

Course Contact Information

Professor

B. Keith Jenkins
Email: jenkins@sipi.usc.edu [please include "EE 660" in the subject line]
Office hours: TBD

Teaching Assistants

Fernando Valladares Monteiro
Email: fvallada@usc.edu
Office hours: TBD

Min Zhang
Email: zhan980@usc.edu
Office hours: TBD

Yao Zhu
Email: yaozhu@usc.edu
Office hours: TBD

All office hours will be conducted by Zoom or Webex. Links will be posted on the course web site.

Graders

Name: Mingxi Lei
Email: mingxile@usc.edu

Name: Pruthvi Sumanth Kakani
Email: pkakani@usc.edu

Your homework assignment will be graded by one of the graders with supervision from one of the TAs and myself. Each week, each grader will grade a different subset of the students' assignments. The initials of the person that graded your assignment will be at the top of your graded assignment or the comment field; this is the person to contact if you have questions about the scoring on your homework. If you don't think the outcome of that discussion is fair or correct, then contact the TA in charge of grading for that assignment. And, of course you may contact the professor if you want a final judgment.

Course Outline

Number of lectures per topic is approximate.

Introduction

1. Course introduction [Murphy] {1 lecture}

   Administrative information; introduction to the course and to machine learning

2. Key issues and concepts in machine learning. {1 lecture}

Regression

3. Multidimensional regression [Murphy] {3 lectures}

   Linear regression, maximum-likelihood and MAP estimation, ridge regression, Bayesian regression. Learning linear and nonlinear relationships.

4. Logistic regression [Murphy] {1 lecture}

Foundations of learning: Bayesian

5. Bayesian concept learning {1 lecture}

Foundations of learning: complexity

6. Feasibility of learning [AML] {1.5 lectures}

   Deterministic and statistical views; Hoeffding inequality (for bounding expected error on unlabeled data); inductive bias (model or data assumptions; e.g., parametric models, local smoothness)

7. Complexity of learning 1: generalization; estimation of error on new data; implications in dataset usage [AML] {3 lectures}

   Generalization bound, effective number of hypotheses, VC dimension, model complexity, sample complexity, dataset methodologies

8. Complexity of learning 2 [AML] {1.5 lectures}

   Bias-variance decomposition, learning curves, overfitting

Foundations and methods of learning: managing and controlling complexity

9. Regularization part 1 [AML] {1 lecture}

   Regularization as soft order constraints

10. Model selection [AML and Murphy] {1 lecture}

    Model selection and validation; consequences on generalization error bounds

11. Regularization part 2; feature reduction; sparsity [Murphy] {2 lectures}

    Bayesian and MAP estimation for feature reduction; quadratic regularization; l_1 regularization, lasso, and sparsity; comparison of l_1 and l_2 regularizers; nonconvex regularizers and l_0 regularization; bridge regression*

12. Principles and pitfalls of learning [AML] {1 lecture}

    Occam's Razor, Axiom of Non-Falsifiability, Sampling Bias, Data Snooping

Graphical and nonlinear methods of learning

13. Boosting techniques and decision trees [Murphy] {3 lectures}

    Adaptive basis models; classification and regression trees (CART); random forests; boosting (Adaboost).

Semi-supervised and unsupervised learning methods

14. Semi-supervised learning for classification [Zhu] {3 lectures}

    Overview, including inductive vs. transductive semi-supervised learning; mixture models and Expectation Maximization for semi-supervised learning.

15. Unsupervised learning for clustering: statistical techniques [Xu] {1 lecture}

    Statistical techniques including mixture models; Maximum Likelihood; Expectation Maximization

16. Unsupervised learning for clustering: other techniques [Murphy and Xu] {2 lectures}

    Similarity measures; evaluating clustering quality and choosing K; hierarchical and graph clustering (agglomerative, divisive, Bayesian)*

Other topics*

17. Optional selected topic(s) of student interest {~1 lecture}

* *As time permits.*

Student work and grading

1. Homework assignments

   There will be approximately one assignment per week. Assignments will generally include some pencil-and-paper problems, some computer problems, and some reading.

   All homework assignments will be posted on D2L on the day assigned. Your solutions will be turned in by uploading pdf file(s) to the D2L assignment dropbox, by the due date and time: one pdf file of your answers to the homework problems, and one pdf file of your code if the assignment included computer problems. Note that the code file must be computer readable (not scanned and not a screen shot).

2. Final Project

   The final course project is a medium-sized project, and will be in the second half of the semester. You will choose your own dataset and problem to work on, subject to certain conditions and guidelines that will be given in the project assignment. Your project will typically involve design, analysis, and demonstration of a machine learning system to solve your chosen problem. Only work done specifically for this class project will count in the project grade. This project will take place approximately during Weeks 10-14, inclusive. Written final project reports and your code will be due near the end of the semester (near the last day of classes, or possibly on the day of the scheduled final exam, to be decided); exact day and time to be announced in the project assignment.

   Your Final Project will be graded by the following criteria: inclusion of required elements; understanding and interpretation (of approach, algorithms used, and results); technical soundness and final performance; quantity and quality of effort; code clarity and organization; and report write-up (clarity, conciseness, and completeness).

3. Possible Midterm Project (TBD)

   The Midterm Project will be a smaller project based on the AML book material. It will be similar to an extended homework assignment. You will work with primarily synthetic data, on problems that are assigned. You will simulate various aspects of learning theory and will assess probabilistic outcomes by running a computer experiment multiple times and collecting statistics of the results. You will compare your numerical results with theory. It will show in depth work and understanding of machine learning topics such as generalization error, statistical sampling, overfitting and underfitting, VC dimension, bias and variance, and regularization. The Midterm Project, if we decide to have one, will take place approximately during Weeks 6-8, inclusive.

   Your Midterm Project will be graded by the following criteria: correctness and completeness; quality organization of work; clarity of write up.

4. Possible Mid-Semester Quiz (TBD)

   If we decide to have it, its date and time will be announced (would l likely be on or near Tues., Oct. 13). It would be short-answer questions, approximately 30-45 minutes, online only.

5. End-of-Semester Quiz

   Will take place near the last day of classes (near Tues., 11/24/2020), or on the official scheduled final examination day (Thursday, 12/3/2020, as determined by the Viterbi School of Engineering). Will tentatively be short-answer questions, approximately 60-90 minutes, online only.

Grading breakdown

Assignment	% of Grade
Homework	25
Mid-semester quiz or project	15
Final Project	35
End-of-semester quiz	20
Online participation	5
TOTAL	100

Note that the above grading table is subject to some minor changes or refinements near the beginning of the semester.

Policy on Collaboration and Individual Work in this Class

Collaboration on techniques for solving homework assignments and computer problems is allowed, and can be helpful; however, each student is expected to work out, code, and write up his or her own solution. Use of other solutions to homework assignments or computer problems, from any source including other students, before the assignment is turned in, is not permitted.

For class projects, general collaboration to resolve issues, or to clarify technical material, is allowed. Use of internet as well as journal and conference literature is encouraged. However, each student (or team) does their own work and writes up their own report. The author(s) of the report are presenting themselves as having done the work described in the report. Any reported work, explanations, information, or code that is obtained from others must be cited as such; instructions for doing this will be given with the project assignment. Including such work in the report without citing it amounts to plagiarism.

Of course, collaboration on quizzes is not permitted.

Please also see below for additional policies that apply to all USC classes.

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. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, policy.usc.edu/scientific-misconduct.

Support Systems

Student Health Counseling Services - (213) 740-7711 – 24/7 on call
engemannshc.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
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 Services (RSVP) - (213) 740-4900 – 24/7 on call
engemannshc.usc.edu/rsvp
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED) | Title IX - (213) 740-5086
equity.usc.edu, titleix.usc.edu
Information about how to get help or help a survivor of 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.

Bias Assessment Response and Support - (213) 740-2421
studentaffairs.usc.edu/bias-assessment-response-support
Avenue to report incidents of bias, hate crimes, and microaggressions for appropriate investigation and response.
The Office of Disability Services and Programs - (213) 740-0776
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
studentaffairs.usc.edu/ssa
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
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, 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
Non-emergency assistance or information.