EE 660 **Course Syllabus** Fall 2018

V2, 8/24/2018

Machine Learning from Signals: Foundations and Methods

Administrative information

Times and days

Lecture: TuTh 3:30 - 4:50 PM, OHE 122 and DEN@Viterbi

Discussion session: Wednesday 12:00 – 12:50 PM, OHE 122 and DEN@Viterbi

Catalogue description

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

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.

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 Matlab or Python for most homework computer problems. A small number of homework problems may require Matlab only. For the class project, students may use Matlab, Python, and/or C/C+. (To use other languages for the class project, check first with the TA or instructor.) Access to Matlab is provided on campus; off-campus students will need to code and to run Matlab at their location if they don't come to campus. All students will be responsible for installing and maintaining their own Python distribution if they use Python.

A Matlab software package, PMTK, will be used as part of this course. It is freely available from https://github.com/probml/pmtk3 ; follow the Readme download and setup instructions.

For Python software packages, users will typically use numpy, pandas, and sk-learn.

Note that for some computer problems, you may be required to code up the problem yourself, without the use of libraries or software packages; and for portions of your course project you may find it advantageous to do so.

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

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

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

Course web site

courses.uscden.net

This uses the Deisre2Learn (D2L) system. The site includes:

– Course materials (handouts, homework assignments, lecture notes, lecture videos, etc.), which will be posted as we progress through the semester.

– Discussion forum (piazza), which can be accessed from here.

– Course calendar, showing events and deadlines.

– Grade book, showing your scores on assignments to date.

– Turning in assignments, and retrieving graded assignments.

Course Contact information

Professor

B. Keith Jenkins
Email: jenkins@sipi.usc.edu
Office location: EEB 404A
Phone: 213-740-4149
Fax: 213-740-6618
Office hours: Wed. 3:30 – 5:00 PM, Fri. 11:00 AM – 12:30 PM

TA's

Fernando Valladares Monteiro
Email: fvallada@usc.edu
Office location: EEB 322
Office hours:
Mon. 3:30 – 5:00 PM

Mehrdad Kiamari
Email: kiamari@usc.edu
Office location: EEB 512
Office hours:
Thur. 12:00 – 1:30 PM, Fri. 4:00 – 5:30 PM

Graders: TBA

Course Outline

There may be some minor changes in the topics or ordering. 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; feature reduction; sparsity [AML and Murphy] {3 lectures}

   Regularization as soft order constraints; 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, l_0 regularization, and bridge regression

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

    Model selection and validation

Graphical and nonlinear methods of learning

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

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

12. Kernel methods (theory and practice) [Murphy] {0.5-1 lecture}

    Mercer kernels, kernel machines; examples of support vector machine variants.*

Semi-supervised and unsupervised learning methods

13. 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.

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

    Statistical techniques including mixture models; Maximum Likelihood; Expectation Maximization

15. 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*

16. Selected topic(s) of student interest. {1-2 lectures}

    A list of topics will be generated by suggestion and discussion. Topics to be covered will be chosen by discussion and vote.

* *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. Computer project

   Will be primarily in the second half of the semester. Students will choose their own dataset and problem to work on, subject to certain conditions and guidelines that will be given in the project assignment. Only work done specifically for this class project will count in the project grade. Final project reports will be due near the last day of classes; exact day and time to be announced in the project assignment.

3. Midterm exam

   Tuesday, 10/23/2018, 3:30 – 4:50 PM PDT.

4. Final exam

   Tuesday, 12/11/2018, 2:00-4:00 PM PST (per USC official schedule of final exams).

Examination ground rules will be decided by group discussion and vote.

Course grade criteria

Assignment	% of Grade
Homework	22
Midterm	26
Project	26
Final exam	26
TOTAL	100

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" https://policy.usc.edu/student/scampus/part-b. 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.

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, computer problems, or the course project, from any source including other students, before the assignment is turned in, is not permitted. Of course, collaboration on exams is not permitted.

Discrimination, sexual assault, intimate partner violence, stalking, and harassment are prohibited by the university. You are encouraged to report all incidents to the *Office of Equity and Diversity/Title IX Office* http://equity.usc.edu and/or to the *Department of Public Safety* http://dps.usc.edu. This is important for the health and safety of the whole USC community. Faculty and staff must report any information regarding an incident to the Title IX Coordinator who will provide outreach and information to the affected party. The sexual assault resource center webpage http://sarc.usc.edu fully describes reporting options. Relationship and Sexual Violence Services https://engemannshc.usc.edu/rsvp provides 24/7 confidential support.

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://ali.usc.edu, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* http://dsp.usc.edu 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.