USC Viterbi
School of Engineering

Ming Hsieh
Department of Electrical Engineering

Course Announcement

EE 660
Machine Learning from Signals: Foundations and Methods
Fall Semester 2017

The course will:

- Familiarize the student with a variety of effective techniques for learning from signals and data
- Enable the student to understand how to match machine-learning algorithms to a given application domain, and to predict error rate on unseen data
- Describe theoretical underpinnings of machine-learning techniques, and their implications in algorithm behavior and design
- Provide the student with opportunities to apply machine learning techniques to synthetic and real data in various application domains

Topics include:

- Linear and nonlinear techniques for classification (output is a category) and regression (output is a real or integer value)
- Approaches to learning in three realms: supervised (train from labeled samples), unsupervised (train from unlabeled samples), and semi-supervised (train from some labeled samples and some unlabeled samples)
- When is learning feasible and reliable: complexity of learning, generalization, estimation of error on new data
- Techniques for matching learning to the application domain: regularization and overfitting, model selection, approximation-generalization tradeoff
- Feature selection and sparse models
- Other topics of student interest

Prerequisites: EE 441 and EE 503

Recommended preparation: EE 559 or CSCI 567; familiarity with Matlab

Time & Location: Tu Th 3:30-4:50 PM, OHE 136; plus one breakout session per week (TBD)

DEN students: This class will also be offered on the den@Viterbi system.

Grading: Homeworks, course project, two quizzes, one final exam.

Instructor: Prof. B. Keith Jenkins, EEB 404A, jenkins@sipi.usc.edu, (213) 740-4149.

EE 660 | **Course Syllabus** | **Fall 2017**

DRAFT 4/20/2017

Machine Learning from Signals: Foundations and Methods

Administrative information

Times and days:

Lecture: TuTh 3:30 - 4:50 PM, OHE 136 and DEN@Viterbi

Discussion (breakout) session: One 50-minute session per week; to be scheduled

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 the student can learn about many of the plethora of machine learning techniques that now exist, on his or her own as needed.

Prerequisites: EE 441 and EE 503.

Recommended preparation: EE 559 or CSCI 567

Computer Hardware/Software Requirements

Students are required to have familiarity with Matlab, and with a coding language of their choice. 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.

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.

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:

– 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).

– 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 at the beginning of the semester.

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 system. The site will include:

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

– Discussion forums, 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.

For help and technical support with the site, go to:

– Support > Getting Started (or Support > Contact Information)

Course Contact information

Professor

B. Keith Jenkins
Office: EEB 404A
Email: jenkins@sipi.usc.edu
Phone: 213-740-4149
Fax: 213-740-6618
Office hours: MW 3:30-5:00 PM

TA

TBA
Office:
Email:
Phone:
Office hoiurs:

Grader

TBA

Course outline

There may be some minor changes in the topics, primarily in their grouping and ordering. Number of lectures per topic is approximate.

Introduction, course overview, background review

1. Administrative information. Introduction to the course and to machine learning. [Murphy] {1 lecture}
2. Continue introduction. Key issues and concepts in machine learning. {1.5 lectures}

Regression

3. Multidimensional regression [Murphy] {2 lectures}

   Linear regression, maximum-likelihood and MAP estimation, ridge regression, Bayesian regression.

4. Logistic regression [Murphy] {1 lecture}

Foundations of learning: complexity

5. Feasibility of learning [AML] {1 lecture}

   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)

6. Complexity of learning; generalization; estimation of error on new data [AML] {2 lectures}

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

Foundations and methods of learning: managing and controlling complexity

7. Review of convexity and optimization {1 lecture}
8. Regularization and overfitting; comparison of regularizers; sparsity [AML and Murphy] {2 lectures}

   Overfitting; quadratic regularization; l1 regularization, lasso, and sparsity; comparison of l1 and l2 regularizers

9. Sparse models and feature selection [Murphy] {2 lectures}

   Bayesian variable selection, l0 regularization; algorithms for optimization with l0 and l1 regularizers; nonconvex regularizers and bridge regression

10. Boosting techniques and decision trees [Murphy] {2 lectures}

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

11. Model selection; comparing classifiers; approximation-generalization tradeoff [AML and Murphy] {2 lectures}

    Model selection and validation, Bayesian model selection, properties of estimators, bias-variance tradeoff

Nonlinear methods

12. Nonlinear models and representation [AML] {1 lecture}

    Basis function representation; nonlinear mappings; effect on VC dimension.

13. Kernel methods (theory and practice) [Murphy] {2 lectures}

    Examples of kernels (radial basis function, Mercer, pyramid match); kernel machines; kernel trick. Support vector machine variants.

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] {1.5 lectures}

    Statistical techniques including mixture models; Maximum Likelihood; Expectation Maximization

16. Unsupervised learning for clustering: other techniques [Murphy and Xu] {2.5 lectures}

    Similarity measures, evaluating clustering quality; hierarchical and graph clustering (agglomerative, divisive, Bayesian); criterion function methods

Other topics (if time permits)

17. Selected topic(s) of student interest. {1.5 lectures}

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

Student work and grading

1. Homework assignments

   Will include some pencil-and-paper problems, some computer problems, and some reading. Some computer problems will use Matlab; others will involve the student writing code in his or her language of choice.

2. Computer project

   Will be in the second half of the semester.

3. Two quizzes

   Will be spaced during the semester. Each quiz will be scheduled and announced in advance, and will be approximately 30 minutes in duration, closed book. Each quiz will consist of short-answer questions.

4. One final exam

   Will be held on Tuesday, 12/12/2017, 2:00-4:00 PM (per official schedule of classes and exams).

Course grade criteria:

Assignment	**% of Grade**
Homework	25
Quizzes	20
Project	25
Final exam	30
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 homeworks, computer problems, or computer projects, from any source including other students, before the assignment is turned in, is not permitted. Of course, collaboration on exams or quizzes 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.