LEARNING FROM DATA

Instructor: Prof. Prakash Ishwar (office: PHO 440, tel: 358-3499, e-mail: pi@bu.edu)

Classes: T,R, 3:30-5:15 (PHO 202)
Office hours: 5:45-7:15 pm on Thursdays (PHO 404/428) and Tuesdays (PHO 442)

Description:
This is an introductory course in statistical learning covering the basic theory, algorithms, and applications. This course will focus on the following major classes of supervised and unsupervised learning problems: classification, regression, density estimation, clustering, and dimensionality reduction. Generative and discriminative data models and associated learning algorithms of parametric and non-parametric varieties will be studied within both frequentist and Bayesian settings in a unified way. A variety of contemporary applications will be explored through homework assignments and a project.

Prerequisites:
Solid foundation in Probability, e.g., EC381 or EK500 or EC505, Linear Algebra, e.g., EK102 or MA142, Multivariate Calculus, e.g., MA225; prior experience with Matlab, e.g., EK 127 is important. Good computer programming skills, e.g., EC327, is desirable.

Outline:

- **Introduction**: overview, key concepts and methodology, probability and linear algebra review, optimal decision rules
- **Classification**: Gaussian discriminant analysis (linear and quadratic), nearest neighbor classifiers, naive Bayes and Bayesian naive Bayes, logistic regression, support vector machines (SVMs), kernel trick, multi-class algorithms, decision trees
- **Regression**: linear (ordinary) least squares, robust linear, ridge, kernel, Lasso, trees
- **Density Estimation**: GMM and the EM algorithm, kernel methods
- **Clustering**: k-means/medoids and connection to EM for GMMs, spectral clustering, hierarchical clustering
- **Dimensionality Reduction**: PCA, kernel PCA, MDS, Isomap, LLE, Laplacian eigenmaps,

Grading:

- **25% Assignments** ~9 combination of Matlab and analytical exercises;
  Due: 4:55pm of a weekday; Worth: 100 points each;
  Late submission penalty: 5 points per minute.
- **30% Project** Team project involving algorithm development; Presentation dates: Dec. 10, 6:30–8:30pm and Dec. 11, 5:30–8:30pm.
  Details to follow later in the semester.
- **40% Quizzes** 3 quizzes 45 minutes each.
  Tentative dates: Oct. 2, Nov. 1, Nov. 27
- **5% Class-participation** For constructive and proactive engagement during class, office hours, and on Piazza.
Course web site: http://learn.bu.edu for registered students only. Will contain lecture slides, notes, links, discussion materials, and other useful information related to the course.

Course references: There is no formal textbook. There is no single textbook that covers the material that I want to introduce in this course in the manner that I want to. Therefore, you will need to rely on lectures and supplementary material that will be uploaded regularly to the course web site. Below is the list of books that can prove useful for various parts of this course should you like to explore. Each book is on reserve in the Science and Engineering Library (max. 24 hour check-out period).


Matlab: Each computer assignment will involve the use of Matlab in order to illustrate and compare the main algorithms discussed in the lectures. You are encouraged to use workstations in PHO 305. Registered students should contact the lab administrator enghelp@bu.edu for an account (if you do not already have one – please check first) and apply for card access to PHO305 (again, only if you do not already have access – please check first) via Zaius at:

http://www.bu.edu/dbin/eng/zaius/

PHO305 is available to registered students on Tuesdays 11am-1pm and Fridays 12pm-2pm. Information about how you can access Matlab remotely can be found here:

http://collaborate.bu.edu/engit/RemoteMatlab

Copyright notice: The copyright on all ENG EC 503 course materials, including lecture slides/notes, assignments, solutions and solution code is asserted by Professor Prakash Ishwar and the specific individuals identified by him. All rights are reserved. Other than for personal use by registered students, the ENG EC 503 course materials or any portion thereof may not be reproduced or used in any manner whatsoever (including posting on a public forum like github) without the express written permission of Professor Prakash Ishwar. Acting otherwise would go against the ethical code of conduct expected of students at Boston University and from a legal standpoint it may even constitute a violation of copyright law.

Academic integrity, plagiarism: Collaboration is essential for the course project, permitted on homeworks, but illegal in exams. In project reports, you may paraphrase relevant ideas from references, but not quote sentences verbatim from them. Homework collaboration = only discussion. If you allow your solution or code to be viewed by anyone or reversely you see someone’s solution or code then you have gone beyond collaboration. You may discuss problems, but you must create a solution by yourself. If there is collaboration in a homework, all collaborators must be explicitly acknowledged and the type and extent of collaboration must be clearly explained. Each collaborator must turn in his/her individual analysis/code and description of results. All solutions (including code) will be automatically checked for plagiarism against solutions from all registered students, solutions from previous semesters, and also solutions available from online sources. The student handbook defines academic misconduct as follows: “Academic misconduct occurs when a student intentionally misrepresents his or her academic accomplishments or impedes other students’ chances of being judged fairly for their academic work. Knowingly allowing others to represent your work as theirs is as serious an offense as submitting another’s work as your own.” Please see the student handbook for procedures that will follow should academic misconduct be discovered.