ENG EC503: Introduction to Learning from Data

Course Information

Lecture Time and Place: EPC 209, Tu/Th 3:30-5:15.

Recitations/Discussion Sessions/GTF Office hours:

1. Monday, 5:00-6:30, PHO 404/428 (**Office hours**).
2. Friday, 6:30-8:00, PHO 205 (**Recitation**).

Lecturers:

Francesco Orabona, 8 St. Mary's St., Room PHO 430, tel: 617-353-1542,
fo@bu.edu, `http://francesco.orabona.com/`.
Office hours: Tu/Thu: 1:30-2:30.
Yannis Paschalidis, 8 St. Mary's St., Room PHO 429, tel: 617-353-0434,
yannisp@bu.edu, `http://sites.bu.edu/paschalidis`.
Office hours: Tu/Thu: 12:00-1:00.

We would recommend to try Piazza first (see below) for getting answers to well-formulated questions. The best way to reach us is via e-mail; we can arrange a meeting time outside regular office hours.

Graduate Teaching Fellow:

Qianqian Ma, maqq@bu.edu.

M.S. Teaching Fellows:

Boran Hao, brhao@bu.edu,
Yu Zhao, yuzhao95@bu.edu.

Description: This is an introductory course in machine learning covering the basic theory, algorithms, and applications. The course surveys a variety of topics, covering both supervised and unsupervised learning problems, and including classification, regression, kernels, robustness and regularization, concepts from learning theory, clustering, dimensionality reduction, generative models, neural networks, and deep learning. A variety of contemporary applications will be explored through homework assignments and a term project.

Course Websites:

Website:	`http://learn.bu.edu/`
Discussion board:	`http://piazza.com/` Signup Link: `http://piazza.com/bu/spring2019/ec503` Class Link: `https://piazza.com/bu/spring2019/ec503/home`

Piazza: We will be using Piazza as a discussion board. You have all been registered and you should have received an invitation to join. The system is highly catered to getting you help quickly and efficiently from both the course staff and your fellow classmates. Rather than emailing questions, we encourage you to post your questions on Piazza.

Prerequisites: Probability, e.g., EC381 or EK500 or EC505, Linear Algebra, e.g., EK102 or MA142, Multivariate Calculus, e.g., MA225, and some level of mathematical maturity. Prior experience with Matlab, e.g., EK 127 is important. Good computer programming skills, e.g., EC327, is desirable.

Textbooks: The required textbook for the class is:
[UML] Shalev-Shwartz and Ben-David. Understanding Machine Learning: From Theory to Algorithms (Cambridge University Press, 2014) The book is available online at
`http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning`.

Supplemental recommended books:
[CML] Daumé. A Course in Machine Learning.
`http://www.freetechbooks.com/a-course-in-machine-learning-t905.html`.

[CO] Boyd and Vandenberghe. Convex Optimization.
`http://web.stanford.edu/~boyd/cvxbook/`.

Grading: There will be regular homework assignments, two in-class exams, and a term project. Your grade will be formed as follows:

1. 15% Homework. We will count the best $n-1$ scores out of n assignments.
2. (25+25=)50% In-class exams.
3. 30% Term project.
4. 5% Attendance and class participation **only if** your overall homework score exceeds 85%.

Attendance: You will find that active class attendance and compilation of class notes are essential in this course. We will be posting slides but we will also use the blackboard; it will be your

responsibility to take notes. Because the topics we will cover build upon each other, if you fall behind you may find that you are lost and not able to follow the lectures.

Homework: Homeworks will be assigned regularly; a bit less frequently than on a weekly basis. They will be due one week after the date issued. We will only accept hard copies of your homework during class, either handwritten or typed (except for some computational assignments for which special instructions will be given). Homework submission by email will not be accepted. Deadlines will be strictly enforced. Although homeworks represent only 15% of the grade you will find that they are *essential* to the learning process. We strongly encourage you to work on them independently. Often, it is easy to follow another person's solution but much harder to come up with your own. Past experience has shown that the performance in the class is highly correlated with your ability to solve problem sets on your own! We will offer homework help at the discussion meeting times and office hours.

Rules of Conduct: You *may* collaborate in study groups on the solution of homeworks. An *acceptable* form of collaboration is to discuss with others possible approaches for solving the problems and then fill the details and write your solutions independently. Copying the solution that someone else has written is *unacceptable* and at times transparent. If you do collaborate you *should* acknowledge your collaborators in the write-up for each problem. We view this as essential!

Needless to say that we expect students to adhere to basic, common sense concepts of academic honesty; presenting another's work as your own or cheating on exams will not be tolerated. Knowingly allowing others to represent your work as their own is as serious an offense as submitting another's work as your own. BU takes academic integrity very seriously. More information on BU's Academic Conduct Code, with examples, may be found at `http://www.bu.edu/academics/policies/academic-conduct-code`.

Make-up Exams: There will be no make-up exams. If there is a legitimate reason for missing an exam, then the scores of other exams will be used appropriately to compensate for the missed exam. If there is no legitimate reason provided for missing an exam, a grade of zero will be assigned for the missed exam.

Term project: In lieu of a final, you will have to complete a project applying some of the knowledge you have acquired in this course. You will present your project in a brief oral (or poster) presentation and submit a written final report. The report should be typed and concise; you should use your judgment as to how much is too much or too little.

You will have to work in small groups of 4-5 people.

There are many alternatives for the project. We want you to take the responsibility and specify the topic; you should view this more as a research task rather than as a homework problem. We expect that before Spring break, you will formulate a concrete proposal for what you plan to do. You may get in touch with us if you want to discuss it.

Projects can be of different types. Here is a partial list:

- **A case study**. You can choose an application area that you are interested in, identify a specific problem, locate a relevant dataset, and solve the problem by implementing and comparing a variety of methods. In such a project the emphasis will be on exploiting the special structure of the particular problem, devise or use appropriate methods, and experiment computationally to obtain insight and maximize effectiveness/performance of the methods.
- **Original theoretical research**. You may investigate any topic theoretically and try to extend the available theory (or create your own theory). Clearly, this can be risky as there are usually no guarantees that you will get results. Still, we would like to see your new theory or method applied to a dataset. You may consider choosing something that has a case study as a fallback option.
- **A Survey paper**. (In this case the number of persons in the team should dictate the number of papers you survey). Select a set of papers on some topic related to the course and write a critical and insightful survey report. You should demonstrate in-depth understanding of the chosen subject. Some of the information your report should provide includes:
    1. An exposition of the problem, why it is interesting, applicability, important issues.
    2. An exposition of the key assumptions.
    3. A description of the main results, their importance, intuitive explanation, proof techniques if they are interesting, sensitivity of the results to the assumptions.
    4. Application of the method to a dataset.
    5. Suggestions for future research directions, possible extensions and variations.

Copyright notice: The copyright on all EC 503 course materials, including lecture slides/notes, assignments, solutions and solution code is asserted by the instructors. All rights are reserved. Other than for personal use by registered students, the 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 the instructors. Acting otherwise would go against the ethical code of conduct expected of students at Boston University and from a legal standpoint it may constitute a violation of copyright law.

Incomplete grades: Incomplete grades will not be given to students who wish to improve their grade by taking the course in a subsequent semester. An incomplete grade may be given for medical reasons if a doctor's note is provided. The purpose of an incomplete grade is to allow a student *who has essentially completed the course* and who has a legitimate interruption in the course, to complete the remaining material in another semester. Students will not be given an opportunity to improve their grades by doing extra work.

Drop dates: Students are responsible for being aware of the drop dates for the current semester. Drop forms will not be back-dated.

Tentative Syllabus.

1. Introduction and logistics.
2. Review of Linear Algebra and Probability.
3. Review of Optimization (UML, Secs. 12.1, 14.1, 14.2, 14.4).
4. PAC learning (UML, Chapters 2, 3).
5. Agnostic PAC learning (UML, Chapter 4).
6. Linear Regression (UML, Chapter 9).
7. Bias-Variance Trade-off and Model Selection.
8. Perceptron.
9. Support Vector Machines (UML Secs. 15.1, 15.2, 12.3).
10. Duality and kernels (UML Secs. 15.3, 15.4, 15.5, 16.1, 16.2).
11. Regularization (UML Secs. 13.1, 25.1.3, 26.3, 26.4).
12. Stochastic Gradient Descent.
13. Multiclass Classification (UML Secs. 17.1, 17.2, 17.3).
14. Nearest Neighborhood and Decision Trees (UML Secs. 18.1, 18.2, 19.1, 19.3).
15. Boosting.
16. Clustering.
17. Dimensionality Reduction (UML Secs. 23.1, 23.2).
18. Generative Models.
19. Neural Networks and Back-propagation.
20. Deep Learning Overview.
21. Applications of deep learning.
22. Online Learning.