Boston University, College of Engineering
Department of Electrical and Computer Engineering

ENG EC503 (Ishwar) – Fall 2024

LEARNING FROM DATA

Instructor: Prof. Prakash Ishwar (e-mail: `pi@bu.edu`)
Teaching Assistant: Param Budhraja (e-mail: `paramb@bu.edu`)
Grader: Luying Ruan (e-mail: `ruanly@bu.edu`)

Lectures: Tue + Thr, 1:30-3:15 pm (PHO 205)
Discussions: Mon, 6:30-8:15 pm (EPC 204)
Instructor's office hours: Tue + Thr 6:00-7:30 pm (PHO 428)
GST's office hours: Wed + Fri 4:30-6:00 pm (PHO 428)

Description:
This is an introductory graduate course in (classical) machine learning covering the basic principles and methods of four major non-sequential supervised and unsupervised learning problems namely, classification, regression, clustering, and dimensionality reduction. A variety of contemporary applications will be explored through homeworks and a project.

Prerequisites:
Probability (EK381 or equivalent), Linear Algebra (EK102 or equivalent), Multivariate Calculus (MA225 or equivalent), *Matlab* (EK125 or equivalent).

Syllabus:

- *Key concepts, terms, and technical tools*: training, testing, cross-validation, performance-evaluation, under- and over-fitting, inner-product, norm, orthogonal projection, eigen-decomposition, SVD, empirical first- and second-order statistics, gradient, Hessian, optimization, convex set, convex function, subgradient, stochastic sub-gradient descent.
- *Nearest-Neighbor (NN) methods*: k-NN classification and regression, k-centers clustering and vector-quantization.
- *Linear methods*: *Classification:* artificial neuron, Fisher-LDA, logistic-loss, single-layer feed-forward artificial neural network (ANN), support-vector machine (SVM); *Regression:* ordinary least squares (OLS), ridge-regression, support-vector regression (SVR);
  Dimensionality-reduction: principal-component analysis (PCA).
- *Kernel-methods*: Representer-theorem, kernelization of learning and inference, kernel versions of kNN, SVM, ridge-regression, k-means, PCA.
- *Multilayer feed-forward ANNs*: architecture, universal-approximation, backpropagation algorithm;
- *Selected Topics (as time permits)*: decision trees and random forests, boosting, document classification with bag-of-words model, robust-regression, sparse-regression, spectral clustering, nonlinear dimensionality reduction methods, probabilistic framework for learning and inference.

Grading:

22%	Homeworks	9 HWs consisting of *Matlab* and/or analytical exercises. Will take better of "all 9 HWs" and "best 8 of 9 HWs". Due: 10:55pm of due date (upload report to Gradescope + code to Blackboard). Early submission bonus: 5% (capped at max score) Late submission penalty: 5% per hour after first hour.
50%	Exams	2 exams. Dates (tentative): Oct. 17 and Dec. 2 in person in PHO 205 and EPC 204, respectively.
25%	Project	Team project (3-4 members) involving algorithm development. Details to follow later in the semester. Presentation: Dec. 9 (6:30–8:15pm) in EPC 204 and Dec 10 (1:30–3:15pm) in PHO 205.
3%	Class-participation	For attendance and constructive and proactive engagement during lectures, discussions, office hours, and on Piazza.

Outcomes: As an outcome of successfully completing this course, students will:

- understand basic theoretical principles and algorithmic methods of learning from data,
- be able to use computer-based machine learning tools,
- know how to analyze and extract information from real-world data,
- be able to select and optimize appropriate machine learning methods and tools for various real-world problems,
- be able to understand and communicate key ideas from articles and technology related to machine learning.

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.

References: This course has no textbook. You will need to rely on lectures, discussions, office hours, and course materials that will be uploaded regularly to the course web site. Below is a list of reference books which you may consult should you like to explore further. Each book is on reserve in the Science and Engineering Library (max. 2 hour check-out period; also available electronically online).

- K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.
- T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: data mining, inference, & prediction*. Springer, 2nd edition - 2009.
- C.M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*. Wiley-Interscience, 2nd edition - 2000.

Matlab: Each computer assignment will involve the use of *Matlab* to illustrate and compare the main methods discussed in the lectures and discussions. You could run *Matlab* on your own computer using BU's site-wide licensing program. See:

`http://www.bu.edu/tech/services/cccs/desktop/distribution/mathsci/matlab/`

Alternatively, you could run *Matlab* remotely on BU machines. See:

`https://www.bu.edu/engit/knowledge-base/citrix/citrix-how-to/`

You may use your own computers to complete the homeworks and/or project. You may also use the workstations in PHO 307, 6:30-8:30pm Mon–Fri. Contact the lab administrator `enghelp@bu.edu`

for an account and apply for card-access using Zaius: `http://www.bu.edu/dbin/eng/zaius/`

Copyright notice: © The copyright on all ENG EC 503 course materials, including lecture slides/notes, discussions, homeworks, 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.

Audio-visual recordings during class: Instructor consent is needed before recording during class due to copyrighted materials. Such recordings must not be posted on any public forum without instructor consent.

Academic integrity, plagiarism: Collaboration is essential for the course project, permitted on homeworks, but illegal in exams.
Homework collaboration = only discussion. If you allow your solution, answers, or code to be viewed by anyone or reversely you see someone's solution, answers, 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 nature and extent of collaboration must be clearly explained. Each collaborator must turn in their individual analysis/code and description of results.**
You cannot quote verbatim from others or AI-generated results (even if you acknowledge them). In project reports, you may *paraphrase* relevant ideas from references, but not quote sentences verbatim from them.
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 their 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.

Generative AI tools: are useful as a learning aid and for "text-smoothing" if you are not a native speaker of English, but beware that you can get wrong answers or/and wrong explanations. For the long-run it is better to *train yourself* to generate correct solutions by seeking guidance from the course instructors. **If you use AI tools, then you must acknowledge it in your work.**

Inclusion: I consider the classroom to be a place of learning where all individuals are expected to contribute to provide a respectful, welcoming and inclusive environment for every member of the class irrespective of how they identify themselves.

Disability accommodations: If you are a student with a documented disability or believe you might have a disability that requires accommodations, requests for accommodations must be made in a timely fashion to Disability & Access Services, 25 Buick St, Suite 300, Boston, MA 02215; 617-353-3658 (Voice/TTY). Students seeking academic accommodations must submit appropriate medical documentation and comply with the established policies and procedures `https://www.bu.edu/disability/`

BU health-safety guidelines and policies: BU strongly recommends staying up-to-date with vaccines, wearing a high-quality mask in crowded indoor spaces, and testing if you are feeling unwell, to reduce the risk of disease transmission, especially for individuals who have a higher risk of severe illness. Students should be compliant with all BU health-safety policies. For BU health-safety guidelines and policies, please visit:
`https://www.bu.edu/chiefhealthoffice/`

Illness/Medical issues: These will be resolved by case-specific discussions with the instructor. Student must inform the instructor by email as soon as possible and be prepared to provide suitable documentation. This may also require contacting BU's disability access services.