EECE 5644: Introduction to Machine Learning and Pattern Recognition

# 2021 Fall Term

**Classes:** Mon-Wed at 14:50-16:30 Live/recorded Zoom meetings & in class meeting at WVG108

**Email Group:** [eece5644@googlegroups.com](mailto:eece5644@googlegroups.com) Use for general questions, comments, so Q&A benefits everyone.

Please send clarification requests on all assignment questions to the google group for everyone’s benefit.

**Prof. Deniz Erdogmus**

E-mail: [erdogmus@ece.neu.edu](mailto:erdogmus@ece.neu.edu) Use for personal inquiries and please always include [EECE5644] in the subject line.

Office: Live Zoom office periods will be announced via google calendar, and you will receive reminder emails as well.

**TA:** Bruna Girvent Email: [girvent@ece.neu.edu](mailto:girvent@ece.neu.edu) Office: In video (Zoom or similar), to be announced via Google Calendar.

**Course Objectives:** Machine learning is the study and design of algorithms that enables computers/machines to learn from experience/data. This course is an introductory course on machine learning covering a range of algorithms, focusing on the underlying models behind each approach, to enable students to learn where and how to apply machine learning algorithms and why they work. The course also emphasizes the foundations to prepare students for research in machine learning. The subjects covered include: Bayes decision theory, maximum likelihood parameter estimation, model selection, mixture density estimation, probabilistic graphical models, support vector machines, neural networks, decision trees, feature selection and dimensionality reduction, ensemble methods: boosting and bagging.

## Prerequisites: Probability EECE 3468/MATH3081 or equivalent for undergraduates, EECE 7204/DS5020 or equivalent for graduate students, knowledge of linear algebra

## Programming Requirement: Must be a self-sufficient programmer (Matlab, Python, C/C++, R are commonly used by students.)

## Suggested Textbooks: A textbook is not required, but I recommend getting one as a formal reference.

Kevin P. Murphy, *Machine Learning: A Probabilistic Perspective*,MIT Press 2012. {Top choice}

Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer 2006.

R. O. Duda, P. E. Hart, D. Stork, *Pattern Classification, 2nd Ed*, Wiley and Sons, 2001

T. Hastie, R. Tibshirani, J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*,Springer, 2001.

**Grading:** There will be 4 take-home assignments and one project (required for PhD students, optional for everyone else). The highest 3 of 4 take-home scores will account for 75% of the total score (3x25%) and the larger of the lowest take-home score and the project score will account for the remaining 25%. You can think of the project as an opportunity to improve your total score, possibly leading to a better letter grade. Individual projects are highly discouraged. Teams of 2-3 people are highly preferred.

## Topics That Will Be Covered

Review of linear algebra, probability theory, numerical optimization (including stochastic gradient parameter updates)

Introduction to machine learning: PCA & Fisher LDA

Bayesian decision theory (classification): Expected risk minimization, MAP and ML classification

Bayesian estimation theory (model fitting): Expected loss minimization, MAP and ML parameter estimation

ML parameter estimation algorithms: EM algorithm and its application to GMMs, MLE for logistic in regression/classification

Model (order) selection: Hyperparameter optimization, Bayesian Information Criterion, Cross-validation (bootstrap, K-fold)

Neural networks: multiplayer perceptrons (including CNNs), recurrent NNs (including LSTMs)

Support vector machines, support vector regression

Kernel regression, Gaussian Process Modeling (with kernel machines), Kernel PCA, Kernel LDA, Kernel density estimation (KDE)

Clustering: k-means, hierarchical, mean-shift (with KDE), spectral (with kernel machines)

Combining classifiers: bagging, boosting (ADABOOST), decision trees

Feature dimension reduction, independent component analysis

**Frequently Asked Questions**

**Assignments:** All take home assignments will require you to do a mixture of math and programming. You will submit results for questions designed to illustrate how concepts we cover in class are used in practice on data. Your submissions will include appropriate amounts of mathematical derivation, code (in an appendix or online repo you point us to), and numerical results tabulated and visualized appropriately in order for you to make a compelling case that convinces the grader you understood the material and used the methods appropriately in the specified problem setting. You will have some guidance on what kind of numerical and visual results to submit, but you will also need to think about what constitutes effective visualization or numerical summary of your computational experiments. Over the term assignments will require less mathematical derivation and more programming, but programming will always be a significant component of the assignment, so you must be able to write code from scratch and be self-sufficient in figuring out how to use your selected programming environment. Take-home assignments will be due within 7-10 days of dissemination, as indicated for each assignment. You will upload your answers (math, numerical results, code) to Canvas by the deadline. **Late submissions are not accepted and deadlines are not extended.** If you have a health or other extraordinary circumstance, let Deniz know immediately to make an alternative individual plan.

Project deliverables are due before the presentations in class. Depending on the number of teams, we will use the last one or two lecture times for team presentations. Deliverables for the project are report, presentation, code, data… Everyone in the team gets the same(usually generous) project score, so it is important for teams to ensure each member contributes approximately equally. Anyone who does not participate in a team project will be assumed to have opte for grading option (b) specified above. Project topics are unrestricted. The only requirement is to identify a problem of interest and tackle it appropriately with machine learning theory and methods. Rigorous methodology is more important than how good the results turn out to be.

**Score-to-letter grade mapping:** A is guaranteed for 90 and above. From this threshold going down by 5-point steps, I obtain default thresholds for lower letter grades. However, these thresholds are slightly adjusted downward if justified by the score distribution, so that large gaps in consecutive sorted scores in class are where letter grade changes happen. This prevents students from missing the one higher letter grade by a relatively small score difference from the next threshold. The threshold adjustments always benefit the student compared to the default thresholds computed using 5-point decrements.

**Theory versus practice:** The course emphasizes rigorous machine learning practice based on strong theoretical fundamentals and understanding of methods. You will learn the mathematical foundations of all techniques you learn to use. Most of your learning will take place while you work on the assignments (as it should be). The office periods are strongly suggested as supplements to classes. We cover extra material in office periods driven by questions from students, and typically focusing on more practical aspects, to the degree of debugging people’s codes. Consequently, attending office periods is strongly recommended. While in class we need to spend time covering material, during office periods, we are less restricted, hence we will address specific questions from every individual to ensure all attendees understand the material covered and become effective in using machine learning techniques.

**Logistics of Remote Operations**

**Canvas:** We will use Canvas initially for announcements until everyone is receiving emails from the google group. Afterwards, we will only use the assignment submission feature of Canvas to get timestamps on submissions. Please do not rely on Canvas messaging or other features, because we will not be monitoring those. Use the group email for clarification requests or comments related to assignments and my personal email for your personal issues.

**Zoom:** Make sure you have access to a computer with microphone/camera that can join Zoom video meetings. Be ready to share your screen when asking a question, so we can see your math or code as a visual complement to your question in a live session such as an office period. We will frequently investigate your math or your code issues as seen on your screen. Please do not send emails with code or math attachments to ask questions. These are best handled live in Zoom meetings.

**Consent to Video Recordings:** Please note that all live zoom sessions (classes and office periods) of Professor Erdogmus (or designated substitute) will be recorded, since the discussion will be generally related to material covered in this course, including clarification questions on assignments. These recordings will be shared with all class members, and by participating in a live Zoom session, you are consenting to these recordings and their dissemination. In the event of a personal question, we will wait until everyone leaves, then stop the recording, and proceed on Zoom privately. However, the student has to request this private Zoom meeting from Prof. Erdogmus. TA zoom sessions will not be recorded, as they are expected to contain private discussions regarding individual assignment scoring reassessments. Please arrange private meetings with the TA within her declared office periods.

**Google:** We will use a shared gdrive folder for all class materials, google calendar event invitations to control access to class and office periods live on Zoom, and google group for all email communications. You do not need to access the google group forum online. It is disabled and will not be used. All calendar events will send reminder emails to all guests a few minutes in advance as a redundancy measure, if you don’t use google calendar invitations to manage your personal calendar. You do not need a google account, but if you do have one that you would like to use for this class, please either define your @northeastern email as an alternate email for that google account, or request membership to the google group using your @gmail address. Associating your google group member email with a google account will give you the benefits of coordination across google services. When requesting membership with a @gmail address, please indicate your @northeastern email in the “reason” box, so I can verify you are registered for class.