Machine learning EE 645

Narayana Santhanam

Spring 2025

Welcome to EE645!
This course is a graduate level course in machine learning.

Pre-requisites The Spring 2025 edition welcomes students from multiple departments: the traditional STEM departments of ECE, other engineering, Computer Science, Astronomy, Math and others, but also from other departments such as (but not only limited to) economics, business and urban planning. The course centers around a few topics that are carefully selected, but which give students an opportunity to spin off and pursue an angle that is best matched with their background and research agendas.

Basic familiarity with statistics, some probability and linear algebra, and programming is required, supplemented with a willingness to learn fundamentals. Ideally, students should have had some exposure to using machine learning/AI (even if it is only YouTube videos or playing with them as black boxes). The goal is for all students to pick up the central tenets of this field in a way that can guide them to use AI/machine learning in a sophisticated and nuanced fashion. You can take this course even if you have taken other ICS courses on the topic, or if you have taken the large language models course in the ECE department.

This course will also have *mentors*, students who have taken prior editions of the course, and we hope you will agree to mentor students down the line.

Material The course will be divided into several modules. There is no single text for the course, but there are several excellent resources that may serve you well beyond the course. Here is an (incomplete) list of some of the books we will use, almost all have legal online versions.

- "Dive into Deep Learning": online book by Zhang et. al.
- "Understanding machine learning: From Theory to Algorithms" by Shai Shalev-Shwartz and Shai Ben-David
- "Patterns, Predictions and Actions" by Moritz Hardt and Benjamin Recht
- "Learning theory from first principles" by Francis Bach
- "Neural Network Learning: Theoretical Foundation" by Martin Anthony and Peter Bartlett (an old but really good book)

The following is a brief outline of the modules we will cover. Please note that this is subject to extensive change and will be modified based on class. All modules that are long enough have mini-projects. Not all mini-projects are mandatory, but at least 2 are.

Soft start, Weeks 1-2 We begin with topics you may already be familiar with. This will refresh some fundamentals, set notation, and provide some context to things we will learn, though the aim is not exactly to teach the material again.

If you are not already familiar with them, you will work with me and some mentors to will help you catch up. Topics covered include (i) linear regression with brief mentions of ridge regression and LASSO (ii) neural networks and basic architectures (convolution networks, autoencoders) and stochastic gradient descent, (iii) linear classification (Fisher discriminant), (iv) linear projections, (v) principal components analysis and singular value decomposition. We will also cover basic evaluation metrics for predictions and regression problems, including notions of receiver operating curves, precision/recall curves, sensitivity and specificity as well as other terms commonly used in medical and other domain analysis.

Kernel methods Weeks 3-5 Kernel methods tend to be state of the art when the amount of data is limited. We will look at kernel methods in general, including support vector machines, kernel PCA and Gaussian processes. We will round this topic up with the concept of random Fourier features. There will be a mini-project on this topic.

Regularization, Week 6 ℓ_1 and ℓ_2 regularization, the Matrix norm, residual connections in neural networks, early stopping, constant step sizes and other common techniques in neural networks.

Large Language models and Transformers, Weeks 7-9 Topic models, Non-negative Matrix factorization, attention and the Transformer architecture, Large Language models, in-context learning and applications of Transformers to non-language tasks. There will be a mini-project in using some of the architectures in this section.

Theory of Learning, Weeks 10-13 This covers VC dimension, PAC learnability, Rademacher complexity and kernel methods, as well as PAC Bayes methods and self certified neural networks. This is a math intensive section, and you can work on theoretical problems as a mini-project as well.

Reinforcement Learning, Week 14-15 Dynamic Programming and reinforcement learning, basic procedures and deep RL. You are welcome to attend related VIP project meetings on reinforcement learning for board games, time tbd. It is not mandatory to attend any of the VIP project meetings however. There will be a mini-project here, but the project itself will be introduced earlier in the course than in week 13.

Projects There will be no midterms, but you will have projects to work on for the duration of the course. The preferred language is python, using pytorch for implementing neural networks. You will also get access to koa for the course. Other variations in python for neural networks (keras, for example) or even other languages (R, Matlab if possible) may be acceptable. The caveat is that they will perhaps be much harder for you.

In addition to the mini-projects above, you will also do a longer term project of your choice in teams. We will also have presentations that we will do in the finals week, based partly on the project your team has done, and partly on a choice of topic you will discuss with me.

Office hours Office hours (tentatively): Mon 10:30-11:30am.

Accommodations We welcome everyone and will support you in your learning journey. If you need any accommodation (kokua, child care issues, disabilities, or special circumstances), please let me know and we will try our best to work it out.