GRS MA 752- Mathematical Foundations of Deep Learning

Instructor: Konstantinos Spiliopoulos
Office: 665 Commonwealth Ave, CCDS 438
Email: kspiliop@math.bu.edu
Course web-page: http://math.bu.edu/people/kspiliop/Spring2025MA752.html
Meets: Spring 2025, Tuesday-Thursday 11:00-12:15 at 685-725 Comm Ave CAS 235
Office hours: TBA

Deep learning is a relatively recent field and mathematical work on deep learning, which is the focus of this class, is even more recent. Key mathematical results will be presented in class and often will be assigned as readings. The material of the class will be based on notes that will be provided to the participants of the class. Some recommended readings are

- Lecture notes by Matus Telgarsky: https://mjt.cs.illinois.edu/dlt/
- I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, The MIT Press, Cambridge, MA, 2016.

Course Description:

This course is a rigorous introduction to the mathematical foundations of deep learning. In particular, the course will introduce the students to the theory of universal approximation, stochastic gradient type of optimizers and their convergence properties, statistical learning bounds, approximation theory, depth separation results, neural tangent kernel, mean field overparametrized regime, implicit regularization and noisy dynamics, different deep neural network architectures and their properties, reinforcement learning, Q-learning, several computational aspects such as back propagation, batch normalization and dropout, deep learning for dynamical systems and variational methods.

The course material will be based on theory, methods (both theoretical and computational) and examples from various scientific disciplines. The class will focus on supervised learning.

Course Prerequisites: Introduction to probability and/or stochastic processes (MA 581 or MA583 or equivalents), Differential Equations (MA226 or MA231 or equivalent), Linear algebra (MA242 or equivalent), basic statistical theory and basic programming skills. Some exposure to real analysis at the undergraduate level will also be useful. PDE's, graduate level probability and statistics will be helpful but not necessary. Students are expected to have the knowledge equivalent to undergraduate level probability or stochastic processes, basic statistical theory, linear algebra, differential equations as well as basic coding skills (ideally in Python).

Tentative Course Syllabus:

Note: I reserve the right to change the course syllabus depending on the progress of the class.

- Week 1: Introduction & Overview

  What is deep learning? Some ideas, applications and examples. Overview of course material and course requirements. Review of tasks in machine learning: supervised learning (regression and classification), unsupervised (such as clustering) and reinforcement learning.

- Week 2: Concepts of overfitting and generalization. Linear regression as motivation for neural networks and perceptron.
- Week 3: Review of basic results in probability theory and stochastic processes

  Review of basic notions in probability theory, such as expected value, variance, conditional expectation, stochastic processes, Markov property, Brownian motion, martingales and basic inequalities, concentration inequalities and different modes of convergence. Reproducing kernel Hilbert spaces.
- Week 4: Shallow and feed forward deep neural networks, universal approximation theorems and their proofs, depth separation results, Barron space and derivation of related bounds.
- Week 5: Gradient flow and gradient descent: conditions for convergence to critical points and minimizers.

  Stochastic gradient descent: Robbins-Monro theory and learning rate decay. Effect of regularization. Backpropagation.
- Week 6: Loss landscapes in deep learning: Over-parametrization effects.

  Variants of stochastic gradient descent: SGD with momentum and optimality for Nesterov's accelerated gradient descent. RMSprop and ADAM optimizers. Convergence properties.
- Week 7: Optimization in shallow neural networks- the linear regime. Neural Tangent Kernel: derivation, convergence theory and properties.
- Week 8: NO class, BU spring break.
- Week 9: Nonlinear overparametrization regime: mean field regime, convergence theory and optimality. Comparisons between the linear and the nonlinear regime. Transitioning from discrete to continuum models.
- Week 10: Deep neural network architectures: recurrent neural networks, LSTMs, highway networks, convolutional neural networks, ResNets. Definitions, properties and usage.
- Week 11: Practical considerations when training neural networks: Backpropagation, exploding and vanishing gradients, different initialization schemes and their effects, batch-normalization, dropout.
- Week 12: (Deep) Reinforcement learning and Q-learning. Definition, applications and convergence properties. Asymptotic analysis of reinforcement learning with neural networks.
- Week 13: Deep learning and dynamical systems: solving ODEs and PDEs with neural networks, closure models based on neural networks, convergence properties, effect of architectures, applications areas (finance, statistics, etc).
- Week 14: Variational Methods: estimating densities, evidence lower bound, generative adversarial network (GAN), Wasserstein GAN (WGAN).

- Week 15: Presentation of class projects.

Assigned work and grading: Registered students will be expected to do regular paper reading, complete a few sets of homework problems (approximately every 2-3 weeks) (75%) and present a final project (25%). Needless to say, you should work on the homework on your own, unless otherwise instructed by me. Late homeworks will not be accepted. The grading policy may change depending on the progress of the class.

Please Note: Students are responsible for knowing the Boston University Academic Conduct Code which is posted at

`https://www.bu.edu/academics/policies/academic-conduct-code/`