Teaching Statement

Osbert Bastani

My goal in teaching is to cover material in a way that conveys how the students might have discovered the material themselves. This approach means conveying both the broader context in which the problem is interesting, as well as the core intuition or picture behind algorithms and proofs. I strive to ensure that these intuitions are accessible to all students from diverse backgrounds, and work hard to ensure that students are comfortable asking questions.

At Penn, I have taught several iterations of the introductory applied machine learning class, which is targeted at both advanced undergraduates (as CIS 4190) and masters/Ph.D. students (as CIS 5190); in particular, I have significantly updated lectures to keep pace with the rapidly changing state of the field. I have also designed and taught (with Rajeev Alur) a new seminar course on trustworthy machine learning (a CIS 7000 course), emphasizing important concepts such as robustness, fairness, and interpretability, which are important for deploying machine learning systems in real-world settings. Finally, I have mentored a number of undergraduate and masters students, several of whom have published first-author papers and have gone on to do Ph.D.s.

Teaching applied machine learning (CIS 4190/5190). The applied machine learning course at Penn has two sections, one for advanced undergraduates (CIS 4190) and one for masters and Ph.D. students (CIS 5190). It provides students with an introduction to the key concepts in machine learning necessary to train and deploy machine learning models such as linear regression, random forests, and deep neural networks. One major challenge in teaching machine learning is that there is a wide variety of different algorithms that emerged from disparate lines of research, including computer science, statistics, neuroscience, and optimization. As a consequence, the material can often feel like a hodgepodge of different techniques without any unifying theme. To alleviate this difficulty, I reorganized the course to relate each algorithm to a central idea from learning theory—namely, that the key decisions when designing a learning algorithm are the model family (or hypothesis class), loss function, and optimization algorithm. I adjusted the lectures to emphasize these decisions across all algorithms in the course, pointing out when algorithms might not fit neatly into this framework for historical reasons. These changes help make machine learning feel like a unified subject, and makes it easier to reason about the differences between different algorithms.

Another challenge is that machine learning is a rapidly changing field. I strove to keep the lectures up to date with these changes. First, I expanded the lectures on deep learning, which are central to understanding modern machine learning techniques. These lectures emphasize the compositionality of neural networks, as well as the internal workings of popular backpropagation frameworks such as PyTorch. Along the same lines, I added lectures on trustworthy machine learning (specifically, topics such as robustness, interpretability, and uncertainty quantification), and on generative AI (including transformers, large language models, and prompt engineering).

Finally, due to the widespread applicability of modern machine learning, the class caters to a wide range of students from different disciplines. Thus, it is crucial to ensure that the material is accessible to students from a broad range of backgrounds. To this end, I provide a significant amount of resources for students on both the mathematical background and on programming in Python, including lecture notes, Python worksheets that step them through key concepts, and recitations. The first homework is designed to help students practice this material and ensure that they have the necessary background.

Teaching trustworthy machine learning (seminar class). I recently designed an advanced class on trustworthy machine learning with Rajeev Alur. Existing machine learning classes in the curriculum focus on the basic concepts and algorithms, which ignores a number of significant challenges to training and deploying machine learning models in practice. For applications where machine learning can have significant impact on peoples' lives, we often require the machine learning model to satisfy certain desiderata beyond just optimizing model performance. In particular, the class covers four of the most common desirable properties: robustness (both distributional and adversarial), uncertainty quantification, fairness, and interpretability. In addition, it emphasizes connections between the techniques used in different parts of the course—for instance, abstract interpretation appears in several contexts, including verifying robustness, uncertainty quantification, and fairness verification. The seminar version of this course was a pilot, and we are planning on offering this course regularly as part of the curriculum for the new B.S.E. in Artificial Intelligence.

Research advising. I mentor students on research at all levels (undergraduate, masters, and Ph.D.). I have advised (or am advising) nine Ph.D., seven masters, and eight undergraduate students at Penn, in addition to working closely with several others at Penn and elsewhere. I have graduated two Ph.D. students; one is now a professor at POSTECH and another is starting a Postdoctoral position at Microsoft Research. I have advised two masters theses and one undergraduate thesis; also, two masters students and four undergraduate students have published papers in top-tier conferences under my supervision. Three masters and one undergraduate students working with me have gone on to Ph.D.s at Stanford, Columbia, NYU, and Penn.

For Ph.D. students, my goal is to guide them to think critically about research so they can learn to formulate and solve their own problems. For them to become a mature researcher, I believe they must develop their own taste in problems, which is challenging since it must be both personal (not just following the latest trend) but also of interest to others (others must be excited about their direction). Typically, I initially provide them with concrete problems to work on. As they progress, I encourage them to explore their own interests, while giving critical feedback to ensure they are working on interesting problems. As a consequence, my students have worked on a wide range of problems depending on their interests: one has worked on incorporating structure into deep generative models, with a focus on music and poetry generation; one has worked on foundation models for robotics control, which have demonstrated impressive results on real robots; and one has worked on leveraging lambda calculus to expose parallelism in neurosymbolic programs, especially chaining large language models. For undergraduate and masters students, my goal is to provide them with exposure to what it is like to work on research problems—I provide them with concrete directions in their area of interest, and work with them to find a manageable project and make progress on each step of the research process.

I aim to have a diverse group, and work hard to recruit students accordingly. I have graduated one female Ph.D. student (who is starting a Postdoctoral position at Microsoft Research), am currently advising one, and have one more joining in Fall 2024. I have advised three female masters students (including two theses and one who went to Columbia for her Ph.D.) and four female undergraduates (including one who went to Stanford for her Ph.D.).