JAY PUJARA – TEACHING STATEMENT

Teaching and mentoring matter. I find nurturing students' discovery of machine learning invigorating. I constantly strive to make sure students are learning – challenging them to master new concepts and apply them to real problems while reflecting on innovative ways to engage them with the material. My comprehensive experience includes teaching undergraduate and graduate courses, managing a large mentoring load, which currently includes twelve PhD and MS students, and outreach spanning innovative classes for high school students to tutorials at top conferences for machine learning experts.

Teaching

In Winter 2016, I was the sole instructor for TIM 245, a graduate-level **Data Mining** course offered through Technology and Information Management engineering department at UC Santa Cruz, with an enrollment of 25 students. The target audience of this course was CS/CE Masters students and PhDs from non-CS disciplines. I made a significant effort to write new homeworks, projects, quizzes and exams that emphasized the **real-world skills students need** to succeed in academia or industry. **Students loved my class**: participation in the evaluations and my score (3.9/5) were well above average. For weeks afterward, their friends would stop me in the hallway to ask if I was teaching again soon. Several of my students asked to continue working with me after the quarter ended, resulting in collaborations I discuss in the next section.

In addition to teaching at the graduate level, I have served as a teaching assistant for CMSC 421, an advanced undergraduate **Artificial Intelligence** course for CS students at the University of Maryland in Fall 2011, with an enrollment of 40 students. In addition to explaining difficult concepts during office hours, helping students each week with applied projects, and writing and grading assignments and exams, I developed several **lectures and in-class exercises on AI and Game Playing**. While TAing, I helped introduce a completely new curriculum for the course, centered on AI assignments for building agents to play PacMan. **Students really enjoyed connecting lectures in class with a fun, tangible problem**. One innovation that really motivated students to get creative with their final projects was a leaderboard system I implemented to run nightly competitions between agents and post the results the next day. Students would vy for the top spot, and at the final class meeting we had a ceremony that involved PacMan-themed crowns and chocolate for the best teams.

As a professor, I plan to teach both undergraduate and graduate courses. At the undergraduate level, I am excited about introducing students to the areas of Data Science, Machine Learning, and Artificial Intelligence. In addition, I am capable of teaching a variety of graduate courses, including core curriculum courses such as Machine Learning, Probabilistic Models, Data Science, or Natural Language Processing. I would also enjoy sharing my research vision by teaching advanced topics courses on Statistical Relational Learning, Information Extraction, and Knowledge Graph Construction.

Course evaluation excerpts from my Data Mining course:

"Jay set making sure students learned the material as a high priority and was very approachable."
"I had a very good understanding of how to use the different models and conduct a semirigorous research on an unseen dataset"
"The quizzes helped me understand the content covered in the class."
"I particularly like the quality of assignments Jay prepared for the class."
"The time he took in making the assignments was apparent and seemed thoughtful. The ipython format was very helpful in learning the course content."
"I'm really happy that we got into Python, that's some very useful realworld experience we gained."

Mentoring

Teaching a large class is an efficient way of transferring knowledge to many, but equally rewarding is working individually with students and developing research ideas. I am currently mentoring twelve students in different stages of their academic career, and two mentees have graduated recently. Having a diverse group of student advisees enables collaborations across many applications, including probabilistic models for understanding cancer, capacity forecasting in datacenters, network security, sustainability, causal modeling for biological domains, mobile device analytics, healthcare informatics, and financial applications.

With my supervisor, Lise Getoor, I work closely with five PhD students, Dhanya Sridhar, Sabina Tomkins, Pigi Kouki, Eriq Augustine, Varun Embar, and four MS students, Shachi Kumar, Nikhil Kini, Hung Ju Chen, and Johnnie Chang. After completing my Data Mining course, several students asked me to act as a research mentor, including Molly Zhang (PhD), and Masters students Ankit Gupta, Vedashree Bagade, Aziz Albalawi, Anirudh Challa, and Stan Thornhill. I served as Masters advisor and committee chair for Stan Thornhill, who successfully submitted his thesis project in June 2016 and graduated. I generally have weekly meetings with students, and have developed a flexible, on-demand scheduling system to preserve my sanity while making sure students get help when they need it.

Mentoring has been a rich experience, exposing me to problems in new domains and forcing me to re-examine my assumptions about existing work. As a postdoc, I have gained a better understanding of the full mentoring life cycle – from helping students identify promising research ideas, guiding them to form hypotheses, debugging code and experimental results together, and iterating to communicate research results clearly and effectively. Frequently, while explaining a difficult concept to a student, I see it from a new perspective, spawning new research ideas. Working with many students and understanding the diverse research issues underlying each has been challenging, and in adapting to this challenge I have become a more efficient and organized collaborator.

Diversity and Outreach

I have a strong track record of encouraging diversity and investing time and energy in outreach. Half of the students I mentor are women (Dhanya Sridhar, Pigi Kouki, Sabina Tomkins, Molly Zhang, Vedashree Bagade, Johnnie Chang, Shachi Kumar), and I am a stalwart supporter of addressing gender disparity in CS departments. My data mining course included eight women, a visiting student from South America, and several students without traditional CS backgrounds, and I created a nurturing environment for this diverse group of students. My teaching experience outside university settings has helped connect students ranging from elementary school children to senior citizens with the computer science developments shaping our world.

Going beyond the ivory tower has provided a new depth to my approach as a teacher. I have pursued teaching opportunities that span teaching 7-year olds how to make simple video games to helping seniors sign up for their first e-mail address. In the former case, I learned a great deal about engaging impatient students by offering a quick initial “a-ha!” moment and progressively introducing new challenges. In the latter, I gained a new appreciation for learning with disabilities: some of my students would grasp the concepts immediately, but would struggle with the fine motor control necessary for manipulating a mouse.

One of the most rewarding teaching experiences I have had involves an annual tradition of volunteering at the National Youth Science Camp (NYSC), a free program for high school students hosted in West Virginia. NYSC selects two delegates from each US state to attend a three-week science-focused camp. I have designed lectures that introduce bright minds who may have lacked access to CS curriculum to deep discrete mathematics concepts and practical machine learning techniques, and always been impressed with the engagement and enthusiasm of this diverse student group. I have also offered several seminars

at NYSC - miniature courses on focused topics - one on Machine Learning and another on Game Theory.

I have also endeavored to share my expertise in knowledge graph construction to colleagues in my discipline. With Sameer Singh and Bhavana Dalvi, I will be presenting a tutorial on Knowledge Graph Construction at AAAI 2017. We plan to adapt and present this tutorial for specialists at different venues, as well as post the materials online. Beyond official teaching employment, I have been invited to present guest lectures at diverse venues that include advanced Machine Learning classes and research groups in industry (e.g., Allen Institute for Artificial Intelligence) and academia (e.g., Max Planck Institut Informatik).

Philosophy

We are a crossroads for computer science education, especially in the disciplines of machine learning and data science: enrollment is skyrocketing, industry hiring is insatiable, new tools and approaches are rapidly introduced, and technological advances are transforming how students learn and teachers teach. Adapting and flourishing in this complex environment requires thoughtfully navigating both student needs and the flood of new material and teaching approaches. **The two central tenets of my teaching philosophy are finding novel strategies to engage students while emphasizing real-world applications of skills they acquire.** Reflecting on my experiences teaching Data Mining, I illustrate the lessons I learned.

A key ingredient to the success of my course was a proactive approach to making the material accessible, engaging, and useful. I completely rewrote the dense, outdated assignments from previous offerings using modern platforms and providing guidance to students. My guiding motivation was supporting **interactive exploration** using iPython/Jupyter notebooks. These offer an integrated interface providing students explanatory text, illustrations, and figures, as well as self-contained code snippets that they can interactively execute, modify, and experiment. **Students appreciate being able to peer beneath the hood of difficult machine learning concepts and understand how they work.** I did not take their understanding for granted, I designed small quizzes that students completed before each lecture. During class, we would go over the quiz problems together, connecting abstract concepts with a problem on the board they helped solve. This also influenced my teaching, since I could **identify where students were struggling before giving a lecture** and spend time where needed.

Despite being an academic at heart, I find it important to underscore that what students learn is not simply academic. I created assignments for data that mattered and had a clear real-world impact. My students diagnosed airplane crashes using FAA records from data.gov, they delved into medical studies on diabetes and heart disease, and analyzed census data to find trends in income. **I want students to leave my classes feeling ready to change the world.**

Of course, there is always room to improve. Many of my students complained my class was too difficult, assignments took too long, and the amount of material covered was too expansive. Given the broad variety of students taking the course (which had no CS prerequisites), I wish I had included more basic examples for those who had weaker CS backgrounds. In my future, I plan to learn from this experience and develop a curriculum that can address the learning goals of both novices and advanced students.

My mentoring philosophy emphasizes responding to the unique needs of each student. In my mentoring interactions, I find students are so excited about research that they become overwhelmed with possibilities (a problem I deeply empathize with). Often channeling their enthusiasm into structured exploration helps them accomplish more. A frequent and rewarding interaction is “For your method to work, what needs to be true?” followed by “Is that true in your data?” I encourage students to identify the problem they solve, why it is difficult, enumerate their assumptions and hypotheses, simplify the problem setting and generate data to assess whether their approach has promise, and then progressively expand the scope of their approach and test on larger datasets.