# Teaching Statement

Stephen Bach

I have prepared for the privilege of teaching through work as a teaching assistant and extensive mentoring. My objectives are to give computer science students a strong command of the foundations of machine learning, teach practical methods to a wider data-science audience, and mentor students on good research process.

**Teaching Interests** In addition to teaching core machine learning courses, I would like to develop a new course on extracting and analyzing structured information like knowledge bases and information networks from unstructured data. In my interactions with industry, including holding office hours and workshops for the Snorkel project, I find that building models for information extraction is one of the most in-demand skills within machine learning, especially in Silicon Valley. This course will connect topics in machine learning, natural language processing, computer vision, and databases. In the undergraduate curriculum, I am excited to teach upper-level classes that introduce students to topics related to my research, such as machine learning, artificial intelligence, and data science, as well as adjacent topics like numeric optimization and database systems. I can also teach a range of introductory computer science courses, from programming to discrete math and statistics.

I am also eager to contribute to the development of an interdisciplinary data science curriculum, which I believe universities will continue to pursue in the coming years. I will contribute by designing courses focused on data analysis that are accessible to students outside of computer science and co-teaching interdisciplinary courses. To do so, I will draw on my experience collaborating with domain experts in biology, medicine, and social science to develop machine learning applications, as described in my papers on weakly supervised machine learning.

**Mentoring** My approach to mentoring is to emphasize research process. When advising on research projects, my first goal is to ensure that students are conducting experiments that will quickly validate their ideas or give information on which direction they should head instead. Reaching this goal requires developing their skills at breaking large projects down into small, testable hypotheses and continually interpreting the results to reassess their direction. I find that these skills are the strongest predictor of a student's success in research.

I have been fortunate to mentor many excellent students with this approach. For example, Golnoosh Farnadi, a visiting Ph.D. student to our lab, was interested in extending the expressivity of probabilistic soft logic, my dissertation work. I mentored her as she developed and published her research, which was recognized with a **Best Student Paper Award** at the 2015 International Conference on Inductive Logic Programming (ILP) [1] and published in extended form in Machine Learning [2]. I also mentored Himabindu Lakkaraju, a Ph.D. student at Stanford, as she worked on interpretable machine learning, which led to our paper at the 2016 Conference on Knowledge Discovery and Data Mining (KDD) [3]. Alex Ratner, another Ph.D. student at Stanford, co-led the development of Snorkel. I mentored him on running a large project that connected machine learning research with application development. We published our resulting paper in the Proceedings of the VLDB Endowment, a top data systems venue [4]. I have also advised numerous master's and undergraduate students at Stanford on a range of projects, including Henry Ehrenberg (weakly supervised machine learning, now at Facebook), Ines Chami (computer vision for document understanding, now applying to Ph.D. programs), and Thomas Lau (tattoo segmentation in images for fighting human trafficking, now a master's student in statistics at Stanford).

## References

[1] G. Farnadi, S. H. Bach, M. Blondeel, M.-F. Moens, L. Getoor, and M. De Cock. Statistical relational learning with soft quantifiers. In *International Conference on Inductive Logic Programming (ILP)*, 2015.

[2] G. Farnadi, S. H. Bach, M. Blondeel, M.-F. Moens, L. Getoor, and M. De Cock. Soft quantification in statistical relational learning. *Machine Learning*, 2017.

[3] H. Lakkaraju, S. H. Bach, and J. Leskovec. Interpretable decision sets: A joint framework for description and prediction. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2016.

[4] A. J. Ratner, S. H. Bach, H. E. Ehrenberg, J. Fries, S. Wu, and C. Ré. Snorkel: Rapid training data creation with weak supervision. *Proceedings of the VLDB Endowment (PVLDB)*, 11(3):269–282, 2017.