Syllabus
LIN 629 Learnability
Fall 2022
TTh 15:00-16:20 – Frey Hall 326

Jeffrey Heinz

August 23, 2022

1 What is this course?

Humans learn language, but little is understood about how this happens. In fact, there is disagreement about what learning means and what kind of behavior a learning algorithm should reasonably exhibit. This course studies, from a computational perspective, definitions of learnability in addition to how anything could learn something like natural language.

We will study computational learning theories including, but not necessarily limited to, *identification in the limit* and *probably approximately correct* learning. We will also discuss a range of machine learning techniques including, but not necessarily limited to, grammatical inference. We will argue over how various results should be understood and we will read papers making some of these arguments. We will learn to understand the tension between theoretical, analytical approaches to learning and benchmarking on specific tasks.

2 Prerequisites

It is highly recommended that students either have a background in mathematics or computer science or have taken Mathematical Methods in Linguistics and Computational Linguistics 2 at Stony Brook University. Students without such a background will struggle and will only be able to contribute partially to some of the broader discussions of the issues.

3 Course Objectives

By the end of this course, students will be able to critically assess work in machine learning as it applies to problems in language learning. Specifically, they will be able to:

1. Understand multiple formal definitions of learning.
2. Understand fundamental issues in developing such definitions.
3. Understand how some algorithms solve some learning problems.

4. Explain advantages and disadvantages of theoretical, analytical approaches to machine learning vis a vis benchmarking on specific tasks.
5. Critically assess computational learning research in the linguistic literature.

Furthermore, students will begin to conduct original research on learning problems as they apply to natural language. Students will:

1. Develop learning problems related to one's own interests in linguistics.
2. Develop theoretical or applied solutions to these problems.

Ideally, by the end of the course, students will be able to conduct original research in grammatical inference and/or computational modeling of language-learning.

4 Office Hours

The instructor's office hours are Monday 2-3pm and Wednesday 11am to 1pm in SBS N237.

5 Course Materials

There is no textbook for the course. I will make readings and notes available online on the course website.

http://jeffreyheinz.net/classes/22F/

Some of the materials we may use are listed below. Students may also present other research subject to my approval. For example, spectral methods for learning finite-state models and statistical relational learning are two current areas of interest not well-represented below.

Mathematical Treatments (Gold)

- E.M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.
- Daniel Osherson, Scott Weinstein, and Michael Stob. *Systems that Learn.* MIT Press, Cambridge, MA, 1986.
- Dana Angluin. Learning regular sets from queries and counterexamples. *Information and Computation*, 75:87–106, 1987.
- Dana Angluin. Identifying languages from stochastic examples. Technical Report 614, Yale University, New Haven, CT, 1988.
- Sanjay Jain, Daniel Osherson, James S. Royer, and Arun Sharma. *Systems That Learn: An Introduction to Learning Theory (Learning, Development and Conceptual Change).* The MIT Press, 2nd edition, 1999.

Mathematical Treatments (PAC)

- M. Anthony and N. Biggs. *Computational Learning Theory.* Cambridge University Press, 1992.
- M.J. Kearns and U.V. Vazirani. *An Introduction to Computational Learning Theory.* MIT Press, Cambridge MA, 1994.

- L.G. Valiant. A theory of the learnable. *Communications of the ACM*, 27:1134–1142, 1984.
- Dana Angluin and Philip Laird. Learning from noisy examples. *Machine Learning*, 2: 343–370, 1988.

Non-mathematical treatment (PAC)

- Leslie Valiant. *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World.* Basic Books, 2013.

Algorithms

- Dana Angluin. Inference of reversible languages. *Journal for the Association of Computing Machinery*, 29(3):741–765, 1982.
- Alexander Clark and Rémi Eyraud. Polynomial identification in the limit of substitutable context-free languages. *Journal of Machine Learning Research*, 8:1725–1745, Aug 2007.
- Colin de la Higuera. *Grammatical Inference: Learning Automata and Grammars.* Cambridge University Press, 2010.
- Jeffrey Heinz. String extension learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 897–906, Uppsala, Sweden, July 2010. Association for Computational Linguistics.
- Jeffrey Heinz, Anna Kasprzik, and Timo Kötzing. Learning with lattice-structured hypothesis spaces. *Theoretical Computer Science*, 457:111–127, October 2012.
- Jane Chandlee, Rémi Eyraud, and Jeffrey Heinz. Learning strictly local subsequential functions. *Transactions of the Association for Computational Linguistics*, 2:491–503, November 2014.
- Jeffrey Heinz, Colin de la Higuera, and Menno van Zaanen. *Grammatical Inference for Computational Linguistics.* Synthesis Lectures on Human Language Technologies. Morgan and Claypool, 2015.
- Jeffrey Heinz and José Sempere, editors. *Topics in Grammatical Inference.* Springer-Verlag Berlin Heidelberg, 2016. ISBN 978-3-662-48395-4.
- Jane Chandlee, Remi Eyraud, Jeffrey Heinz, Adam Jardine, and Jonathan Rawski. Learning with partially ordered representations. In *Proceedings of the 16th Meeting on the Mathematics of Language*, pages 91–101, Toronto, Canada, 18–19 July 2019. Association for Computational Linguistics.
- Dakotah Lambert, Jonathan Rawski, and Jeffrey Heinz. Typology emerges from simplicity in representations and learning. *Journal of Language Modelling*, 9(1):151–194, 2021.

Perspectives

- Martin A. Nowak, Natalia L. Komarova, and Partha Niyogi. Computational and evolutionary aspects of language. *Nature*, 417:611–617, June 2002.
- Jeffrey Heinz and William Idsardi. Sentence and word complexity. *Science*, 333(6040): 295–297, July 2011.
- Jeffrey Heinz and Jason Riggle. Learnability. In Marc van Oostendorp, Colin Ewen, Beth Hume, and Keren Rice, editors, *Blackwell Companion to Phonology*. Wiley-Blackwell, 2011.
- Adam Albright and Bruce Hayes. Learning and learnability in phonology. In John Goldsmith, Jason Riggle, and Alan Yu, editors, *Handbook of Phonological Theory*, pages 661–690. 2011.

- Alexander Clark and Shalom Lappin. *Linguistic Nativism and the Poverty of the Stimulus*. Wiley-Blackwell, 2011.
- Nick Chater, Alexander Clark, John A. Goldsmith, and Amy Perfors. *Empiricism and Language Learnability*. Oxford University Press, 2015.
- Jeffrey Heinz. Computational theories of learning and developmental psycholinguistics. In Jeffrey Lidz, William Synder, and Joe Pater, editors, *The Oxford Handbook of Developmental Linguistics*, chapter 27, pages 633–663. Oxford University Press, Oxford, UK, 2016.
- Joe Pater. Generative linguistics and neural networks at 60: Foundation, friction, and fusion. *Language*, 95(1):e41–e74, 2019. doi:10.1353/lan.2019.0009.
- Jonathan Rawski and Jeffrey Heinz. No free lunch in linguistics or machine learning: Response to Pater. *Language*, 95(1):e125–e135, 2019.
- Jeffrey Heinz and Jonathan Rawski. History of phonology: Learnability. In Elan Dresher and Harry van der Hulst, editors, *Oxford Handbook of the History of Phonology*, chapter 32. Oxford University Press, 2022.

Latest Argument

- Yuan Yang and Steven T. Piantadosi. One model for the learning of language. *PNAS*, 119 (5), 2022. https://doi.org/10.1073/pnas.2021865119.
- Jordan Kodner, Spencer Caplan, and Charles Yang. Another model not for the learning of language. *PNAS*, 119(29), 2022. https://doi.org/10.1073/pnas.2204664119.

6 Grades

There will be occasional written homework assignments. Collectively, these will count for 25% of the final grade.

Each students will present a paper, or part of a paper, during the course of this class. The presentation will count for 25% of the final grade.

Students will complete a project in the course of the semester which relates to learning and learnability. The project proposal should be 300-500 words, explaining the project and how you plan to go about it. **Proposals are submitted to me and returned to you with feedback in a cycle that repeats until I approve them. Project proposals can be initially submitted anytime but must be approved by November 1, 2022.**

Projects can be theoretical or applied. Projects are due December 16. The project counts for 50% of the final grade.

7 University Policies and Services

7.1 Student Accessibility Support Center Statement

If you have a physical, psychological, medical, or learning disability that may impact your course work, please contact the Student Accessibility Support Center, Stony Brook Union Suite 107, (631) 632-6748, or at sasc@stonybrook.edu. They will determine with you what accommodations are necessary and appropriate. All information and documentation is confidential.

Students who require assistance during emergency evacuation are encouraged to discuss their needs with their professors and the Student Accessibility Support Center. For procedures and information go to the following website: https://ehs.stonybrook.edu//programs/fire-safety/emergency-evacuation/evacuation-guide-disabilities and search Fire Safety and Evacuation and Disabilities.

7.2 Academic Integrity Statement

Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty is required to report any suspected instances of academic dishonesty to the Academic Judiciary. Faculty in the Health Sciences Center (School of Health Professions, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures. For more comprehensive information on academic integrity, including categories of academic dishonesty please refer to the academic judiciary website at http://www.stonybrook.edu/commcms/academic_integrity/index.html

7.3 Critical Incident Management

Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of Student Conduct and Community Standards any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn. Faculty in the HSC Schools and the School of Medicine are required to follow their school-specific procedures. Further information about most academic matters can be found in the Undergraduate Bulletin, the Undergraduate Class Schedule, and the Faculty-Employee Handbook.