Algorithms in Society

Boston University CAS CS591-S1

Spring 2020

Adam Smith

Important Links

- Syllabus: https://docs.google.com/document/d/1hER0O7BK4hlXfRk9RVP7X6p2p6DWeb1Zgy3PDI5Uq68/edit?usp=sharing
- Piazza site: https://piazza.com/bu/spring2020/cs591s1
- Lecture slides: https://drive.google.com/drive/folders/1L1AwjOsIYJWFQjcY5VsjABswH7g0_J1x?usp=sharing
- Homework: https://drive.google.com/drive/folders/1Lbla-Af035cFGbM9bkMWe33SQP2OkbPk?usp=sharing
    - Three homework problem sets, focused on differential privacy

Textbooks and Monographs

We will draw on a few sources for textbook-level presentations of certain material:

- C. Dwork and A, Roth. *The Algorithmic Foundations of Differential Privacy*, 2014. https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf
    - Theoretically-focused introduction to basic algorithmic ideas in differential privacy.
- S. Vadhan. The Complexity of Differential Privacy. Monograph, 2016.
- S. Barocas, A. Narayanan, and M. Hardt: Fairness and machine learning: Limitations and Opportunities. (In progress, 2019) https://fairmlbook.org/
    - In-progress book on fairness in ML (mostly binary classification). Coverage is spotty.
- T. Hastie, R. Tibshirani, J. Friedman. *The Elements of Statistical Learning*, 2nd edition (2017) https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12.pdf
    - Introduction to basic concepts of statistical learning. Chapter 2 is a good starting point for people new to the topic.

- S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms*, 2014. http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning
    - Introduction to learning theory focused on convex optimization (and PAC learning)

Lecture Schedule

Day	Date	Num.	Topic	Reading and Notes
Tue	1/21/2020	1	Course Introduction; Privacy as a general topic; Failures of Naive Anonymization.	Overview material: ● Quick overview of ways that recent technology affects privacy: S. Fowler, What We've Learned From Our Privacy Project (So Far), New York Times, July 2019. ● Introduction to fairness concerns in ML: Barocas, Narayanan, and Hardt. Chapter 1 (Introduction) Topic-specific reading: ● A. Narayanan and V. Shmatikov: Myths and Fallacies of "Personally Identifiable Information", CACM 2010 ● A. Korolova: Privacy Violations Using Microtargeted Ads: A Case Study, JPC 2011. ● M. Hansen. New York Times, December 2018. To Reduce Privacy Risks, the Census Plans to Report Less Accurate Data ● Optional reading ○ A. Tockar: Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset, 2014 ○ A. Narayanan and V. Shmatikov: Robust De-anonymization of Large Sparse Datasets, S&P 2008 ○ L. Backstrom et al: Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography, WWW 2007
Thu	1/23/2020	2	Attacks on apparently aggregate summaries and models: Reconstruction attacks.	● Hastie, Tibshirani, Friedman. Chapters 2 to 2.6. ● Section 1-2 from Dwork et al, Exposed! A Survey of Attacks on Private Data, An*nual Review of Statistics and Its Application* (2017). ○ Ok to skip the proofs. ● Optional: ○ S.Garfinkel, J.. Abowd, C. Martindale. Understanding Database Reconstruction Attacks on Public Data. CACM, 2019.
Tue	1/28/2020	3	Attacks on apparently aggregate summaries and	● R. Shokri, M. Stronati, C. Song, V. Shmatikov. Membership Inference Attacks Against Machine Learning Models. IEEE Security & Privacy (Oakland) 2017

			models: tracing (membership inference) attacks.	• Section 3 from Dwork et al, Exposed! A Survey of Attacks on Private Data, An*nual Review of Statistics and Its Application* (2017).
				  ○ Ok to skip the proofs
				• N. Carlini, C. Liu, Ú. Erlingsson, J.Kos, D. Song. The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks, USENIX Security 2019.
				  ○ Watch the video, read Sections 1, 2, 3, 10.1, and 10.2.

Thu	1/30/2020	4	Differential privacy introduction.	• Dwork and Roth, Chapters 1 and 2.
				• Optional:
				  ○ Read S. Vadhan's presentation: Vadhan, section 1 to 1.5.
				  ○ Watch K. Ligett's introduction to differential privacy (and randomized response): https://www.youtube.com/watch?v=Z1MOWGCyW20

Tue	2/4/2020	5	DP: Basic algorithms.	• Dwork and Roth, Chapter 3.1 to 3.3.
				  ○ Randomized response
				  ○ Laplace Mechanism

Thu	2/6/2020	6	Reconstruction attacks	Guest lecture: Aloni Cohen Reading: Cohen and Nissim, 2019, Sections 1 and 2.
Tue	2/11/2020	7	DP: Basic Algorithms	(No new reading)

Thu	2/13/2020	8	DP: Basic Algorithms	• Dwork and Roth, Chapter 3.3 and 3.4
				  ○ Report noisy max
				  ○ Exponential mechanism
				• Optional: J. Dong. Range and Sensitivity of the Exponential Mechanism (blog post, February 2020)

Tue	2/18/2020		NO LECTURE (Monday schedule)	
Thu	2/20/2020	9	DP: Gaussian noise and "strong" composition	
Tue	2/25/2020	10	DP: Gaussian noise and "strong" composition Homework 1 due.	
Thu	2/27/2020	12	Gradient descent	
Tue	3/3/2020	13	Noisy and stochastic gradient descent	

Thu	3/5/2020	14	Local DP protocols	
Tue	3/10/2020		SPRING BREAK	
Thu	3/12/2020		SPRING BREAK	
Tue	3/17/2020	15	Local DP protocols	

Thu	3/19/2020	16	Statistical notions of "fairness"	• Corbett-Davies et al, A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. Washington Post, Oct. 17, 2016.
				• Berk et al: Fairness in Criminal Justice Risk Assessments: The State of the Art. *Sociological Methods & Research*, 2018. DOI 10.1177/0049124118782533
				  ○ Sections 1-6
				• Chouldechova and Roth. The Frontiers of Fairness in Machine Learning. 2018
				  ○ Skim this.
				Optional:
				• Neil and Winship. Methodological Challenges and Opportunities in Testing for Racial Discrimination in Policing. *Annu. Rev. Criminol.* 2019. 2:73–98.
				• Friedler et al. On the (im)possibility of fairness, 2016
				• Chouldechova, Fair prediction with disparate impact: A study of bias in recidivism prediction instruments, https://arxiv.org/abs/1703.00056, 2017
				• Kleinberg, Mullainathan, and Raghavan, Inherent Trade-Offs in the Fair Determination of Risk Scores, http://arxiv.org/abs/1609.05807, 2016

Tue	3/24/2020	17	Lipschitz treatments ("Individual fairness")	• Dwork, Feldman, Hardt, Pitassi, Reingold, Zemel. Fairness Through Awareness, 2011 Optional: • Dawid, *On Individual Risk*, https://arxiv.org/abs/1406.5540, 2014.
Thu	3/26/2020	18	"Fairness", postprocessing, and composition	• Dwork and Ilvento, Fairness Under Composition • Canetti et al., From Soft Classifiers to Hard Decisions: How fair can we be? ○ Sections 1-3 Optional: • Bower et al, Fair Pipelines, 2017 • Dowrk and Ilvento, Individual Fairness in Pipelines, 2020.
Tue	3/31/2020		Finishing up composition	
Thu	4/2/2020	19	Generalization guarantees	• Generalization guarantees and VC-dimension • Yona and Rothblum. [1803.03242] Probably Approximately Metric-Fair Learning. 2018

Tue	4/7/2020	20	Learning a metric from people	• Ilvento. [1906.00250] Metric Learning for Individual Fairness, 2019. See also this talk. • Jung et al., [1905.10660] Eliciting and Enforcing Subjective Individual Fairness, 2019
Thu	4/9/2020	21	Fairness and feedback effects	• Lily Hu and Yiling Chen. A Short-term Intervention for Long-term Fairness in the Labor Market. In Proc. of The Web Conference (WWW), Lyon, France, April 2018. • Lydia T. Liu, Sarah Dean, Esther Rolf, Max Simchowitz, Moritz Hardt. Delayed Impact of Fair Machine Learning ICML 2018. See also this blog post. • Lydia T. Liu, Ashia Wilson, Nika Haghtalab, Adam Tauman Kalai, Christian Borgs, Jennifer Chayes. The Disparate Equilibria of Algorithmic Decision Making when Individuals Invest Rationally. See also this blog post and this video.
Tue	4/14/2020	22	Intersecting groups	• Ursula Hébert-Johnson, Michael Kim, Omer Reingold, Guy Rothblum. Multicalibration: Calibration for the (Computationally-Identifiable) Masses, ICML 2018 • Michael Kearns, Seth Neel, Aaron Roth, Zhiwei Steven Wu. Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness, ICML 2018.
Thu	4/16/2020	23	Fairness—guest talk by Gal Yona	• Michael P. Kim, Aleksandra Korolova, Guy N. Rothblum, Gal Yona. Preference-Informed Fairness . *ITCS* 2020.
Tue	4/21/2020	24	Games, learning,and regret	We will go over some technical tools that come up in many of the fairness papers—no-regret dynamics and zero-sum games.
Thu	4/23/2020	25	Explanations and interpretability	• Z. Lipton. The Mythos of Model Interpretability , 2016 • Wortman-Vaughan and Wallach. A Human-Centered Agenda for Intelligible Machine Learning (PDF). Draft book chapter, January 2020 • Doshi-Velez and Kim, Towards A Rigorous Science of Interpretable Machine Learning, 2017
Tue	4/28/2020	26	Explanations and interpretability	• Kleinberg and Mullainathan. Simplicity Creates Inequity: Implications for Fairness, Stereotypes, and Interpretability 2019
Thu	4/30/2020	27	Wrap up. Adaptive data analysis quick peek.	
Wed	May 6		Project presentations	