Introduction to Differential Privacy

Monday, October 17, 2022 - Friday, October 28, 2022
Workshop for JPSM, University of Maryland

Instructors

Anand D. Sarwate is an Associate Professor in the Department of Electrical and Computer Engineering at Rutgers, the State University of New Jersey, where he has been since 2014. He received a B.S. degree in Electrical Science and Engineering and a B.S. degree in Mathematics from MIT in 2002, an M.S. in Electrical Engineering from UC Berkeley in 2005 and a PhD in Electrical Engineering from UC Berkeley in 2008. From 2008-2011 he was a postdoctoral researcher at the Information Theory and Applications Center at UC San Diego and from 2011-2013 he was a Research Assistant Professor at the Toyota Technological Institute at Chicago, a philanthropically endowed academic computer science institute located on the University of Chicago campus. He was the Online Editor of the IEEE Information Theory Society (2015-2018) and an Associate Editor for the IEEE Transactions on Signal and Information Processing over Networks (2015-2018). Prof. Sarwate received the Rutgers Board of Trustees Research Fellowship for Scholarly Excellence in 2020, the the A. Walter Tyson Assistant Professor Award from the Rutgers School of Engineering in 2018, and an NSF CAREER award in 2015. His interests are in information theory, machine learning, and signal processing, with applications to distributed systems, privacy and security, and biomedical research. He is the author of several papers on differential privacy and differentially private machine learning, and has worked on applying differential privacy in biomedical applications. He has also given tutorial lectures on differential privacy at the Workshop on Information Forensics and Security (WIFS 2014) and the Neural Information Processing Systems (NeurIPS 2017) conference.

Course Objectives

The primary objective of this course is to get participants familiar with the basic conceptual framework of differential privacy, when and where it is applicable, basic techniques for differentially private data analysis, and adapting statistical methodologies to data which has been published under differential privacy. The instructor will motivate the need for rigorous privacy definitions before covering basic definitions for pure, approximate, and local differential privacy as well as basic methods for achieving them such as randomized response, output perturbation and the exponential mechanism. Applications and examples from
research on testing, statistical inference, synthetic data generation, and machine learning will be discussed with pointers to further work. Case studies from industry will illustrate some of the challenges in deploying differential privacy at scale. The live sessions will include application vignettes with code to illustrate and ground the concepts via demonstrations on real data.

**Learning Outcomes:**
After this course, participants should be able to:

- Explain privacy risks from inference attacks.
  - Understand and explain privacy risks from publishing microdata or analyses.
  - Explain how reidentification & reconstruction attacks work.
  - Explain the risks posed by membership attacks & other attacks.
  - Explain how many statistical disclosure mechanisms fail to provide privacy.
- Interpret the definitions of pure, approximate, and local differential privacy.
  - Give simple examples of differentially private strategies.
  - Understand the scope of tasks to which differential privacy is suited.
  - Explain the differences between the threat models and guarantees.
- Apply basic differentially private mechanisms for common statistical tasks.
  - Compute the global sensitivity of basic summary statistics.
  - Apply the Laplace and Gaussian mechanisms to basic operations.
  - Apply the exponential mechanism for differentially private selection.
- Understand how to apply differential privacy to core problems in statistics and machine learning.
  - Explain approaches for hypothesis testing with privacy.
  - Use differential privacy for standard inference methods.
  - Explain challenges in applying differential privacy in small sample settings.
  - Understand how differential privacy is being used for modern machine learning.
- Understand the linear query model approach for differential privacy and its application to synthetic data generation.
  - Describe different approaches to query processing.
  - Describe different approaches for differentially private synthetic data generation.
- Understand challenges arising from implementing differential privacy in practice.
  - Describe mathematical and formal challenges for private information processing.
  - Understand the challenges faced in applying differential privacy to the US Census.
  - Understand the types of decisions differentially private system designers have to make when using differential privacy at scale.
Who Should Attend

Statisticians and statistical analysts in government, universities, business, and nonprofit organizations interested in understanding how differentially private data analysis works and how to approach differentially private data published through instruments such as the 2020 United States Census.

Introductory course work in statistical methods is strongly recommended. Participants should be familiar with descriptive statistics, standard distributions (normal, Laplace, geometric, binomial, etc.), chance selection, expected values, and basic probability calculations. Familiarity with statistical testing (Type I/Type II errors) will also be very helpful.

Course Schedule:

The course will be in an Online Format from Oct 17 to Oct 28, Participants would have online access to the course packet (slides) and to the recorded lectures where is presented in 4 sections Monday - Thursday. (We would have the entire week 1 recordings available online and participants could watch them at their own pace, with a recommended viewing of 1 1/2 hours per day for each of the four days).

Friday 10/21 and 10/28 would be a 1:30 hour live group Online discussion with the instructors from 10:00am to 11:30am EST. Week 2 follows the same schedule as Week 1, (recorded lectures M-Th, live Online group meeting Friday).