Anonymity, privacy

Pierre Senellart Debabrota Basu Hervé Chabanne

7 December 2021

Need for anonymization

- Many datasets include personally identifying information (names, IDs, IPs, etc.) and sensible data (clinical data, sexual orientation, religion, etc.)
- EU General Data Protection Regulation imposes constraints on what can be done with personal data
- Machine learning models often based on personal data, potentially leaking sensitive information
- Many technical challenges in dealing with private data without disclosing private information

Curriculum

- Classes on Friday afternoons:
  - online for the first three sessions (21/01, 28/01, 04/02)
  - at Paris-Dauphine, room A301 on 11/02
  - in Issy-les-Moulineaux from 18/02
- 8 sessions + project defenses
- Topics covered:
  - Basics of data privacy and anonymization
  - Measuring anonymity: k-anonymity, l-diversity, m-closeness
  - A probabilistic framework: differential privacy
  - Differentially private mechanisms
  - Local differential privacy
  - Differential privacy and federated learning
  - Processing data anonymously, homomorphic encryption

Evaluation

One homework (40% of the total grade): take one research paper, summarize it, explain it in your own words, comment on its strengths and limitations

One project (60% of the total grade): take one (or more) research paper, build something cool from it (implement it, improve the algorithm, test it on some interesting dataset, etc.), present it in a defense