Exchange programme Vrije Universiteit Amsterdam

Vrije Universiteit Amsterdam - Exchange programme Vrije Universiteit Amsterdam - 2024-2025

Exchange

Vrije Universiteit Amsterdam offers many English-taught courses in a variety of subjects, ranging from arts & culture and social sciences, neurosciences and computer science, to economics and business administration.

The International Office is responsible for course approval and course registration for exchange students. For details about course registration, requirements, credits, semesters and so on, please visit the exchange programmes webpages.

Differential Privacy and Statistical Inference under Differential Privacy

Course Code	E_EOR3_DP
Credits	6
Period	P5
Course Level	300
Language Of Tuition	English
Faculty	School of Business and Economics
Course Coordinator	dr. R. de Vlaming
Examiner	dr. R. de Vlaming
Teaching Staff	dr. J.N. van Brummelen, dr. R. de Vlaming
Teaching method(s)	Study Group, Lecture

Course Objective

By the end of this course, you

1. can explain the difference between local and global differential privacy (DP);
2. can explain the role of a trusted curator;
3. can explain the difference between a synthetic data release, and an interactive interface in which the analyst can put queries to a trusted curator;
4. can construct a basic synthetic data release;
5. can explain when two databases are adjacent;
6. can describe the general purpose of a random mechanism;
7. can define the privacy loss of observing a certain output of a random mechanism;
8. can explain what a distinguishing event is;
9. can give a Bayesian interpretation of privacy loss;
10. can give the mathematical definition of ε-DP;
11. can explain ε-DP in terms of the worst-case absolute privacy loss;
12. can calculate the sensitivity of a query;
13. can apply the Laplace mechanism to a query, such that the result is ε-DP;
14. can explain the trade-off between statistical efficiency and DP;
15. can explain why the Laplace mechanism is not ε-DP in case

a. the number of returned outputs depends on in the input database, and/or
b. the sensitivity is unbounded due to unbounded attributes;

1. can explain why trimming can

a. deal with unbounded attributes,
b. limit the amount of noise required to reach ε-DP, and
c. cause estimators to become inconsistent;

1. can give examples of other mechanisms that do not yield ε-DP results;
2. can give the mathematical definition of (ε, δ)-DP;
3. can describe δ, informally, in terms of an upper bound on the probability of a distinguishing event;
4. can explain why the more versatile (ε, δ)-DP definition can still offer a high level of privacy protection in a probabilistic sense;
5. can explain why requiring (ε, δ)-DP, instead of ε-DP, enables us to apply mechanisms that induce less noise than the Laplace mechanism;
6. can apply basic theorems such as the composition theorem;
7. can implement various DP mechanisms in Python;
8. can calculate *p*-values and standard errors of DP point estimates using simulations;
9. and can implement machine-learning techniques in a DP-manner using Python.

Course Content

Access to data is key for scientific progress. Yet, data is often sensitive in nature and must, therefore, fulfil certain requirements to ensure privacy. For instance, publicly available data should be arranged such that nobody can trace back the individuals in the dataset.

In this course, you will learn about a solid mathematical framework called*Differential Privacy* (DP), that enables you to quantify the loss in privacy by releasing data or results into the public domain. Moreover, you will learn about algorithms that actually give you control over the amount of privacy that is lost by such a release. These algorithms, in essence, simply add some noise to the data.

Intuitively, it should be clear that such perturbations will affect parameter estimates and their uncertainty (e.g., when estimating the effect of years of schooling on income later in life). Hence, the second objective of this course will be to find ways to estimate parameters consistently, to quantify their standard error, and to perform hypothesis tests correctly, when using such algorithms. In other words, you will learn about *Differential Privacy and Statistical Inference under Differential Privacy*.

Additional Information Teaching Methods

Lectures (4 hours a week) and tutorials (2 hours a week).

Method of Assessment

Written or digital exam and assignments.