Internship project description
Numerical inference of differential privacy guarantees

Jan Ramon

11/2023

1 Motivation and context

Over the last decades, there has been an increasing interest in exploiting data. On the other hand, recently there has also been an increasing awareness of the risks of collecting sensitive data centrally, given the frequency of data leaks, hacking or abuse. INRIA's Magnet team is interested in decentralized privacy-preserving machine learning where the sensitive data remains with the data owners, and machine learning is performed collaboratively by these data owners by participating in collaborative algorithms which (through the use of differential privacy and/or encryption) generate the desired statistical models but prevents sensitive data from being revealed. Important ongoing research projects in the team include the TIP, TRUMPET and FLUTE projects. This internship fits into the larger research program including these projects.

Even with perfect security, the output of an algorithm can still leak information. That is why it is interesting to study statistical privacy notions such as differential privacy. There is a large body of work proving differential privacy guarantees for specific machine learning algorithms. Unfortunately, the composition theorems which are commonly used aren't tight bounds. In reality better privacy is achieved than what can be easily proven symbolically, especially if a machine learning algorithm is complex and involves iterations.

In this internship, we want to explore a different approach, where we don't study differential privacy guarantees using classic symbolic proof techniques, but where we attempt to find out how private a given algorithm is by an automated numeric analysis. The idea is inspired on other existing numeric approaches, e.g., if one can't give a simple closed form expression for an integral, it is common to approximate its value using numeric integration, dividing the interval over which one should integrate into intervals for which an accurate estimate can be obtained.

2 Objectives

The goal of this internship project is to research and validate algorithm(s) to give differential privacy guarantees for machine learning algorithms using numerical techniques.

The end result should consist of (a) an algorithm for providing differential privacy guarantees, (b) theory showing correctness / security of this algorithm, (c) empirical results showing how much better the guarantees are compared to classic symbolic approaches.

Depending on the skills and interest of the student, more emphasis can be put on either objective (b) theory or objective (c) experimentation.

3 Plan

Here is a tentative work plan:

- Getting familiar with differential privacy and numeric techniques (3 weeks)
- Formulating the problem, outlining one or more ideas to solve the problem (3 weeks)
- Proving correctness of the suggested approach(es) (2-8 weeks)
- Implementing a prototype and performing experiments (10-4 weeks)
- Completion of the internship report (2 weeks)

The timing can be adapted according to the personal preferences of the student or the requirements of his school.

4 Environment

The project will be conducted in the INRIA MAGNET team. The student will collaborate and interact with various other collaborators in the team. It is possible the student will be asked to present progress to the team or to partners in ongoing projects. An application for ZRR access will need to be made to the FSD.

5 Requirements

We expect the student has a strong knowledge of basic computer science concepts, e.g., data structures and algorithms, and of statistics / machine learning. The topic may involve challenging theory.