Applying Differential Privacy to Machine Learning: Challenges and Potentials

Franziska Boenisch
Fraunhofer AISEC

With data storage becoming more affordable, different parties collect increasingly large amounts of data about individuals. The data collection allows for data analyses that help to improve software, recommend products, or even make large advances in the medical field through tracking user behavior. At the same time, the concern about the individuals' privacy preservation is growing. Differential Privacy (DP) (Dwork, 2006) offers a solution to the potential conflict of interest between privacy preservation and extensive data analyses. Its goal is to allow meaningful data analyses on a whole population while ensuring a mathematically provable level of privacy for the individual.

DP libraries for general data analyses based on statistical queries (e.g. sum, count, min, max) already exist (Wilson *et al.*, 2019). The underlying methods rely mainly on masking the private data by adding noise to the query results. However, the process in the context of machine learning is not that straightforward. Machine learning models usually condensate a given training dataset with a lot of attributes to a few learned parameters. These parameters, therefore, reflect properties of those underlying data and might, thereby, reveal some private information about it. In this work, I propose an adaptation of the definition of DP to the context of machine learning. Furthermore, I present two different DP linear regression models. Based on the presented models, I investigate three main questions. (1) How much can an individual data record change the learned model parameters? (2) Where should the noise be added (to the model parameters, to the cost function, or to the prediction output)? (3) How does the choice in question (2) influence the trade-off between prediction accuracy and individual privacy?

I found that a naive DP linear regression method that adds noise to the prediction output reaches a much lower accuracy than a more advanced method, like the one presented by Zhang *et al.*, which adds noise to the cost function (Zhang *et al.*, 2012). The reason is that the second method needs a smaller amount of noise to reach the same level of privacy. This suggests that the success of applying DP to machine learning depends mainly on the methods used. The focus of further research should, therefore, lie on developing and improving the existing DP methods for different machine learning algorithms.

References

Cynthia Dwork (2006). Differential Privacy. In *Automata, Languages and Programming*, Michele Bugliesi, Bart Preneel, Vladimiro Sassone

crypto day matters 31 (2019)
Beiträge vom 31. Kryptotag, 17.-18.10.2019, Qualcomm, Berlin.
doi:10.18420/cdm-2019-31-26

& Ingo Wegener, editors, volume 4052, 1–12. Springer Berlin Heidelberg. ISBN 978-3-540-35907-4 978-3-540-35908-1.

Royce J. Wilson, Celia Yuxin Zhang, William Lam, Damien Desfontaines, Daniel Simmons-Marengo & Bryant Gipson (2019). Differentially Private SQL with Bounded User Contribution .

Jun Zhang, Zhenjie Zhang, Xiaokui Xiao, Yin Yang & Marianne Winslett (2012). Functional Mechanism: Regression Analysis under Differential Privacy .