Mach Learn
DOI 10.1007/s10994-013-5404-1

Bounds on the sample complexity for private learning and private data release

Amos Beimel · Hai Brenner · Shiva Prasad Kasiviswanathan · Kobbi Nissim

Received: 3 July 2011 / Accepted: 5 August 2013

Abstract Learning is a task that generalizes many of the analyses that are applied to collections of data, in particular, to collections of sensitive individual information. Hence, it is natural to ask what can be learned while preserving individual privacy. Kasiviswanathan et al. (in SIAM J. Comput., 40(3):793–826, 2011) initiated such a discussion. They formalized the notion of *private learning*, as a combination of PAC learning and differential privacy, and investigated what concept classes can be learned privately. Somewhat surprisingly, they showed that for finite, discrete domains (ignoring time complexity), every PAC learning task could be performed privately with polynomially many labeled examples; in many natural cases this could even be done in polynomial time.

While these results seem to equate non-private and private learning, there is still a significant gap: the sample complexity of (non-private) PAC learning is crisply characterized in terms of the VC-dimension of the concept class, whereas this relationship is lost in the constructions of private learners, which exhibit, generally, a higher sample complexity.

Looking into this gap, we examine several private learning tasks and give tight bounds on their sample complexity. In particular, we show strong separations between sample complexities of proper and improper private learners (such separation does not exist for non-private learners), and between sample complexities of efficient and inefficient proper private

Editor Phil Long.

A. Beimel (✉) · K. Nissim
Dept. of Computer Science, Ben-Gurion University, Beer-Sheva, Israel
e-mail: beimel@cs.bgu.ac.il

K. Nissim
e-mail: kobbi@cs.bgu.ac.il

H. Brenner
Dept. of Mathematics, Ben-Gurion University, Beer-Sheva, Israel
e-mail: haib@bgu.ac.il

S.P. Kasiviswanathan
General Electric Research Center, San Ramon, CA, USA
e-mail: kasivisw@gmail.com

Published online: 18 September 2013