A New Analysis of Differential Privacy's Generalization Guarantees

Abstract:

Many data analysis pipelines are adaptive: the choice of which analysis to run next depends on the outcome of previous analyses. Common examples include variable selection for regression problems and hyper-parameter optimization in large-scale machine learning problems: in both cases, common practice involves repeatedly evaluating a series of models on the same dataset. Unfortunately, this kind of adaptive re-use of data invalidates many traditional methods of avoiding over-fitting and false discovery, and has been blamed in part for the recent flood of non-reproducible findings in the empirical sciences. An exciting line of work beginning with Dwork et al. 2015 establishes the first formal model and first algorithmic results providing a general approach to mitigating the harms of adaptivity, via a connection to the notion of differential privacy. Unfortunately, until now, those results were primarily of information theoretic interest, only beating out the simple approach of gathering fresh data for every computation ("sample-splitting") at the scale of many millions of datapoints.

In this work, we give a new proof of the transfer theorem that any mechanism for answering adaptively chosen statistical queries that is differentially private and sample-accurate is also accurate out-of-sample. Our new proof is elementary and gives structural insights that we expect will be useful elsewhere. We show: 1) that differential privacy ensures that the expectation of any query on the conditional distribution on datasets induced by the transcript of the interaction is close to its expectation on the data distribution, and 2) sample accuracy on its own ensures that any query answer produced by the mechanism is close to the expectation of the query on the conditional distribution. This second claim follows from a thought experiment in which we imagine that the dataset is resampled from the conditional distribution after the mechanism has committed to its answers. The transfer theorem then follows by summing these two bounds, and in particular, avoids the "monitor argument" used to derive high probability bounds in prior work.

An upshot of our new proof technique is that the concrete bounds we obtain are substantially better than the best previously known bounds, even though the improvements are in the constants, rather than the asymptotics (which are known to be tight). As we show, our new bounds outperform the naive "sample-splitting" baseline at dramatically smaller dataset sizes compared to the previous state of the art, bringing techniques from this literature closer to practicality.

Joint work with: Christopher Jung, Seth Neel, Aaron Roth, Saeed Sharifi-Malvajerdi (UPenn), and Moshe Shenfeld (HUJI).

This work appeared at ITCS 2020.