Lattice-Based Methods Surpass Sum-of-Squares in Clustering

Ilias Zadik
Department of Mathematics, Massachusetts Institute of Technology
IZADIK@MIT.EDU

Min Jae Song
Courant Institute of Mathematical Sciences, New York University
MINJAESONG@NYU.EDU

Alexander S. Wein
Algorithms and Randomness Center, Georgia Institute of Technology
AWEIN@CIMS.NYU.EDU

Joan Bruna
Courant Institute of Mathematical Sciences, New York University
Center for Data Science, New York University
Center for Computational Mathematics, Flatiron Institute
BRUNA@CIMS.NYU.EDU

Editors: Po-Ling Loh and Maxim Raginsky

Abstract

In this work we show that for an important case of the canonical clustering task of a \(d\)-dimensional Gaussian mixture with unknown (and possibly degenerate) covariance, a lattice-based polynomial-time method can provably succeed with the statistically-optimal sample complexity of \(d + 1\) samples. This is in contrast with the evidence of “computational hardness” for this task, as suggested by the previously established failure of low-degree methods and the Sum-of-Squares hierarchy to succeed with access to \(\tilde{O}(d^{3/2})\) and \(\tilde{O}(d^2)\) samples, respectively.

Keywords: statistical-to-computational gaps, lattice basis reduction, sum-of-squares lower bounds, average-case complexity

Clustering is a fundamental primitive in unsupervised learning which gives rise to a rich class of computationally-challenging inference tasks. In this work, we focus on the canonical task of clustering \(d\)-dimensional Gaussian mixtures with unknown (and possibly degenerate) covariance. Recent works (Ghosh et al., 2020; Mao and Wein, 2021; Davis et al., 2021) have established lower bounds against the class of low-degree polynomial methods and the sum-of-squares (SoS) hierarchy for recovering the clusters with access to \(\tilde{O}(d^{3/2})\) and \(\tilde{O}(d^2)\) samples, respectively. Prior work on many similar inference tasks portends that such lower bounds strongly suggest the presence of an inherent statistical-to-computational gap for clustering, that is, a parameter regime where the clustering task is statistically possible but no polynomial-time algorithm succeeds.

One special case of the clustering task we consider is equivalent to the problem of finding a planted hypercube vector in an otherwise random subspace. We show that, perhaps surprisingly, this particular clustering model does not exhibit a statistical-to-computational gap, despite the aforementioned low-degree and SoS lower bounds. To achieve this, we give an algorithm based on Lenstra–Lenstra–Lovász (LLL) lattice basis reduction (Lenstra et al., 1982) which achieves the statistically-optimal sample complexity of \(d + 1\) samples, building upon the use of LLL in the seminal papers Lagarias and Odlyzko (1985); Frieze (1986) and in the more recent inference settings (Zadik and Gamarnik, 2018; Gamarnik et al., 2021; Andoni et al., 2017; Song et al., 2021). This result extends the class of problems whose conjectured statistical-to-computational gaps can be “closed” by “brittle” polynomial-time algorithms, highlighting the crucial but subtle role of noise in the onset of statistical-to-computational gaps.

References


