CHRIS LIAW, University of British Columbia
Nearly-Tight Sample Complexity Bounds for Learning Mixtures of Gaussians

Estimating distributions from observed data is a fundamental task in statistics that has been studied for over a century. We consider such a problem where the distribution is a mixture of k Gaussians in $\mathbb{R}^d$. The objective is density estimation: given i.i.d. samples from the (unknown) distribution, produce a distribution whose total variation distance from the unknown distribution is at most ϵ.

We prove that $\tilde{\Theta}(kd^2/\epsilon^2)$ samples are necessary and sufficient for this task, suppressing logarithmic factors. This improves both the known upper bound and lower bound for this problem.