

Optimal Kullback-Leibler Aggregation in Mixture Density Estimation by Maximum Likelihood

jeudi 21 juin 2018 09:30 (30 minutes)

We study the maximum likelihood estimator of density of n independent observations, under the assumption that it is well approximated by a mixture with a large number of components. The main focus is on statistical properties with respect to the Kullback-Leibler loss. We establish risk bounds taking the form of sharp oracle inequalities both in deviation and in expectation. A simple consequence of these bounds is that the maximum likelihood estimator attains the optimal rate $((\log K)/n)^{1/2}$, up to a possible logarithmic correction, in the problem of convex aggregation when the number K of components is larger than $n/2$. More importantly, under the additional assumption that the Gram matrix of the components satisfies the compatibility condition, the obtained oracle inequalities yield the optimal rate in the sparsity scenario. That is, if the weight vector is (nearly) D -sparse, we get the rate $(D \log K)/n$. As a natural complement to our oracle inequalities, we introduce the notion of nearly- D -sparse aggregation and establish matching lower bounds for this type of aggregation.

Summary

We study the maximum likelihood estimator of density of n independent observations, under the assumption that it is well approximated by a mixture with a large number of components. The main focus is on statistical properties with respect to the Kullback-Leibler loss. We establish risk bounds taking the form of sharp oracle inequalities both in deviation and in expectation. A simple consequence of these bounds is that the maximum likelihood estimator attains the optimal rate $((\log K)/n)^{1/2}$, up to a possible logarithmic correction, in the problem of convex aggregation when the number K of components is larger than $n/2$. More importantly, under the additional assumption that the Gram matrix of the components satisfies the compatibility condition, the obtained oracle inequalities yield the optimal rate in the sparsity scenario. That is, if the weight vector is (nearly) D -sparse, we get the rate $(D \log K)/n$. As a natural complement to our oracle inequalities, we introduce the notion of nearly- D -sparse aggregation and establish matching lower bounds for this type of aggregation.

Auteur principal: Prof. DALALYAN, Arnak (ENSAE / CREST)

Co-auteur: Dr SEBBAR, Mehdi

Orateur: Prof. DALALYAN, Arnak (ENSAE / CREST)

Classification de Session: Theory around mixtures