Local Risk Bounds for Statistical Aggregation

Jaouad Mourtada  
CREST, ENSAE, Institut Polytechnique de Paris, France
JAOUAD.MOURTADA@ENSAE.FR

Tomas Vaškevičius  
Institute of Mathematics, EPFL, Switzerland
TOMAS.VASKEVICIUS@EPFL.CH

Nikita Zhivotovskiy  
Department of Statistics, UC Berkeley, USA
ZHIVOTOVSKIY@BERKELEY.EDU

Editors: Gergely Neu and Lorenzo Rosasco

Consider a family of predictors \( \{f_\theta : \theta \in \Theta\} \) indexed by an abstract set \( \Theta \). In the aggregation problem, the aim is to combine the given predictors to achieve predictions nearly as accurate as the best one. Given a prior probability measure \( \pi \) on \( \Theta \), classical global bounds in aggregation take the form

\[
E R(\hat{f}) \leq C_{\text{glob}}(n/c, \pi) := -\frac{c}{n} \log \left( \int_{\Theta} \exp \left( -\frac{n}{c} R(f_\theta) \right) \pi(d\theta) \right),
\]

where \( \hat{f} \) is the output of the aggregation procedure, \( R(f_\theta) \) measures the risk of the predictor \( f_\theta \), \( c \) is a universal constant, and \( n \) is the size of the sample. An important feature of aggregation theory is that bounds of the above form can be obtained without imposing any restrictions on the set \( \Theta \). In particular, this allows to handle non-convex classes of predictors.

Aggregation has been studied in both sequential and statistical contexts. The classical results in both cases feature the same global complexity measure \( C_{\text{glob}}(n/c, \pi) \), despite the sequential aggregation problem being intrinsically harder than its statistical counterpart. In this paper, we revisit and tighten classical results in the theory of aggregation in the statistical setting by replacing the global complexity with a smaller, local one. This local complexity measure can be seen as the global complexity computed with respect to a prior distribution \( \pi_{-\frac{n}{c} R} \) that is more tightly concentrated around functions with low population risk; it can be defined as follows:

\[
C_{\text{loc}}(n/c, \pi) \asymp C_{\text{glob}}(n/c, \pi_{-\frac{n}{c} R}), \quad \text{where} \quad \pi_{-\frac{n}{c} R}(d\theta) = \frac{\exp \left( -\frac{n}{c} R(f_\theta) \right) \pi(d\theta)}{\int_{\Theta} \exp \left( -\frac{n}{c} R(f_\theta') \right) \pi(d\theta')}.
\]

We discuss the nature of this improvement in some special cases, including finite classes and Gaussian process priors on functions. Our approach is inspired by the technique of PAC-Bayesian localization introduced by Catoni (2007).

Among other results, we prove localized versions of the classical bound for the exponential weights estimator due to Leung and Barron (2006) and deviation-optimal bounds for the Q-aggregation estimator. Our bounds refine the results of Dai, Rigollet, and Zhang (2012) in the fixed design setting and the results of Lecué and Rigollet (2014) for random design.

Acknowledgments

J. M. and N. Z. acknowledge the support of the France-Berkeley Fund.

Extended abstract. Full version appears as [arXiv:2306.17151, v1], (Mourtada, Vaškevičius, and Zhivotovskiy, 2023)

References


