Non-asymptotic penalization criteria for mixture-of-experts regression models with Gaussian gating functions

TrungTin NGUYEN

Normandie Univ, UNICAEN, CNRS, Laboratoire de Mathématiques Nicolas Oresme, 14000 Caen, France.

Abstract

Mixture-of-experts (MoE) models are a popular framework for modeling heterogeneity in data, for both regression and classification problems in statistics and machine learning, due to their flexibility and the abundance of statistical estimation and model choice tools. Such flexibility comes from allowing the mixture weights (i.e, the gating functions) in the MoE model to depend on the explanatory variables, along with the experts (i.e, the component densities). This permits the modeling of data arising from more complex data generating processes, compared to the classical finite mixtures and finite mixtures of regression models, whose mixing parameters are independent of the covariates. The use of MoE models in a high-dimensional setting, when the number of explanatory variables can be much larger than the sample size (i.e., $p \gg n$), is challenging from the computational and theoretical points of view, where the literature still lacks results for dealing with the curse of dimensionality, in both the statistical estimation and feature selection problems. We aim at estimating the number of components of this mixture, as well as the complexity of the regression relationship using a penalized maximum likelihood approach. To this end, we provide both a weak oracle inequality and an l_1-oracle inequality for MoE regression models with Gaussian gating functions.