

The question of generalization performance of machine learning algorithms is usually tackled using a worst case approach in statistical learning theory. This leads to robust but somewhat loose bounds. Another approach, common in statistical physics of learning and high-dimensional statistics, relies on exact solutions of benchmark models typically based on random design. This leads to more precise predictions at the cost of stronger assumptions. A natural question that can be asked is whether these benchmark models are realistic, and whether the intuition they provide can be used with real data. We will provide some answers to these questions by discussing exact solutions to generalized linear modelling on structured data, more precisely in the context of generic Gaussian mixtures. Technical insight will be given on the proof methods, in particular on approximate message passing algorithms, which have become an important element of high-dimensional probability.