PAC-Bayes, Rademacher and Descriptional Complexities: Three Sides of the Same Coin

Peter Grunwald

Abstract

The race to explain deep learning's remarkable generalization performance has sparked renewed interest in learning bounds. These come in different flavors, depending on what's on the left and what's on the right. On the left we usually find generalization error or excess risk; on the right we have a KL divergence between prior and posterior (PAC-Bayes bounds) or a version of Rademacher complexity. We offer a new bound that accommodates for all of these at the same time.

On the right-hand side, we feature a novel complexity measure COMP that can be relaxed to both KL divergence and Rademacher complexity, in the latter case retrieving the best known excess risk bounds. Surprisingly, COMP is also the minimax regret in sequential prediction under log-loss in a model class derived from the original model/loss and as such can be related to minimum description length.

As to the left-hand side, we account for the strange phenomenon that existing generalization error bounds are typically much weaker than excess risk bounds, the former involving a \sqrt{KL/n} term whereas the latter can, under fast-rate conditions, be as small as O(KL/n). We provide a novel technique leading to a MAC-Bayesian generalization bound that can also have O(KL/n) on the right.

Joint work with Nishant Mehta (ALT 2019) and Zakaria Mhammedi (NeurIPS 2019 and work in preparation).