Preferent compression for tight generalization bounds

Marco C. Campi MARCO.CAMPI@UNIBS.IT
Università di Brescia, Dipartimento di Ingegneria dell'Informazione.
Via Branze 38, 25123, Brescia, Italy

Simone Garatti SIMONE.GARATTI@POLIMI.IT
Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria.
Piazza L. da Vinci 32, 20133, Milan, Italy

Editor: Simone Vantini, Matteo Fontana, Aldo Solari, Henrik Boström and Lars Carlsson

Abstract

Generalization theories aim to assess the quality of data-driven hypotheses (e.g., classifiers or predictors) without relying on validation datasets. A well-known generalization framework is sample compression. It rests on the idea that being able to compress the initial set of data into a subset of small size, while retaining all the information needed to construct the hypothesis, implies low probability of change of compression after introducing a new data point. In turn, this entails good generalization capabilities because in many schemes inappropriateness of the hypothesis ("inappropriateness" means inability to correctly classify in classification problems or inability to correctly predict in more general schemes) results in a change of compression. Our poster builds on a recent article written by the same authors of this submission titled "Compression, Generalization, and Learning", JLMR 2023. The poster illustrates a new foundational theory for compression-based generalization and presents results that enable control over the probability of change of compression under a broadly applicable preference condition. This leads to unprecedentedly tight finite-sample upper bounds on the probability of inappropriateness (the so-called "statistical risk"). Importantly, our results do not require any a-priori restriction on the size of the compressed set. Furthermore, it is shown that lower bounds are also attainable under a suitable additional condition. These lower and upper bounds rapidly converge on top of each other as the number of data points grows, showing that the size of the compressed set serves as a consistent estimator of the statistical risk. All results are established within a fully agnostic setup that does not assume any prior knowledge on the probability distribution of observations. This makes our findings suitable for hyper-parameter tuning, while also bolstering confidence in observation-driven methodologies.

Keywords: statistical learning theory, distribution-free uncertainty quantification, compression schemes, scenario approach.