

Weak Signals: Machine-Learning Meets Extreme Value Theory

mercredi 3 avril 2024 14:30 (50 minutes)

The angular measure on the unit sphere characterizes the first-order dependence structure of the components of a random vector in extreme regions and is defined in terms of standardized margins. Its statistical recovery is an important step in learning problems involving observations far away from the center. In the common situation that the components of the vector have different distributions, the rank transformation offers a convenient and robust way of standardizing data in order to build an empirical version of the angular measure based on the most extreme observations. However, the study of the sampling distribution of the resulting empirical angular measure is challenging. It is the purpose of this talk to explain how to establish finite-sample bounds for the maximal deviations between the empirical and true angular measures, uniformly over classes of Borel sets of controlled combinatorial complexity. The bounds are valid with high probability and, up to logarithmic factors, scale as the square root of the effective sample size. The bounds are next applied to provide performance guarantees for two statistical learning procedures tailored to extreme regions of the input space and built upon the empirical angular measure: binary classification in extreme regions through empirical risk minimization and unsupervised anomaly detection through minimum-volume sets of the sphere.

Orateur: CLÉMENÇON, Stephan (LTCI, Télécom Paris, Institut Polytechnique de Paris)