Séminaire de Probabilités et Statistique

Mardi 25 mai à 14h00
ZOOM

Guillaume Maillard
Université du Luxembourg

Aggregated Hold-out

Cross-validation is widely used for selecting among a family of learning rules. This paper studies a related method, called aggregated hold-out (Agghoo), which mixes cross-validation with aggregation; Agghoo can also be related to bagging. We provide the first theoretical guarantees on Agghoo, ensuring that one can use it safely: at worse, Agghoo performs like the hold-out, up to a constant factor. For the hold-out, oracle inequalities were known in the case of bounded losses, as in binary classification. The approach can be extended, with appropriate hypotheses, to most classical risk-minimization problems. Under weak hypotheses, we obtain an oracle inequality for the penalty parameter in regularized kernel regression with Lipschitz loss. In all these settings, Agghoo verifies an oracle inequality. However, simulation studies suggest that real performance is often much better than what theory can prove in general. In particular, there is a large gain from aggregation that bounds derived from the hold-out are incapable of capturing. As a result, Aggregated hold-out appears to be competitive with standard cross-validation in practice.