Bayesian variable and group selection approach for high-dimensional data

Sarah Filippi*[1]

[1]Imperial College London – Royaume-Uni

Résumé

Few Bayesian methods for analysing high-dimensional sparse data provide scalable variable selection, effect estimation and uncertainty quantification. Most methods either sacrifice uncertainty quantification by computing maximum a posteriori estimates, or quantify the uncertainty at high (unscalable) computational expense. In this talk, we will focus on two specific problem settings: (1) variable selection for high dimensional sparse survival data and (2) the selection of groups of variables under a generalised linear model. For both settings, we develop an interpretable and scalable model for prediction and variable or group selection. Our method, based on a variational approximation, overcomes the high computational cost of MCMC whilst retaining the useful features, providing excellent point estimates and offering a natural mechanism for variable/group selection via posterior inclusion probabilities. We derive posterior concentration rates for the group sparse Variational Bayes approach, compare our methods against other state-of-the-art Bayesian variable selection methods on simulated data and demonstrate their application for variable and group selection on real biomedical data.

*Intervenant