

Information criteria for structured sparse variable selection

Thursday, 17 August 2017 14:30 (30 minutes)

In contrast to the low dimensional case, variable selection under the assumption of sparsity in high dimensional models is strongly influenced by the effects of false positives.

The effects of false positives are tempered by combining the variable selection with a shrinkage estimator, such as in the lasso, where the selection is realized by minimizing the sum of squared residuals regularized by an ℓ_1 norm of the selected variables. Optimal variable selection is then equivalent to finding the best balance between closeness of fit and regularity, i.e., to optimization of the regularization parameter with respect to an information criterion such as Mallows's C_p or AIC. For use in this optimization procedure, the lasso regularization is found to be too tolerant towards false positives, leading to a considerable overestimation of the model size. Using an ℓ_0 regularization instead requires careful consideration of the false positives, as they have a major impact on the optimal regularization parameter. As the framework of the classical linear model has been analysed in previous work, the current paper concentrates on structured models and, more specifically, on grouped variables. Although the imposed structure in the selected models can be understood to somehow reduce the effect of false positives, we observe a qualitatively similar behavior as in the unstructured linear model.

Primary author: Mr MARQUIS, Bastien (Université Libre de Bruxelles)

Co-author: Mr JANSEN, Maarten (Université Libre de Bruxelles)

Presenter: Mr MARQUIS, Bastien (Université Libre de Bruxelles)