

# Supervised feature selection with structure learning for sparse and weak data in high dimensional setting

Annika Tillander<sup>1</sup> and Tatjana Pavlenko<sup>2</sup>

<sup>1</sup>*Linköpings University, Sweden*

<sup>2</sup>*KTH Royal Institute of Technology, Stockholm, Sweden*

## Abstract

In a high-dimensional setting where the informative features are few and the information as such is weak, it is difficult to successfully differentiate between classes. By identifying the relation between features the information can be enhanced. A known dependence structure between features enables block-diagonal approximation of the inverse covariance matrix, then block-wise information strength can be estimated and selection of blocks of features instead of single features. Using only informative blocks in an additive classifier has shown improved classification accuracy [1]. Hence the need of learning the structure of the covariance matrix. A novel block-identifying procedure is suggested and applied to simulated data as well as real data.

## Keywords

Feature selection, High dimensionality, Supervised classification, Structure learning.

## References

- [1] Pavlenko, T., Björkström, A. and Tillander, A. (2012). Covariance structure approximation via gLasso in high-dimensional supervised classification. *Journal of applied statistics* 39, 1643–1666.