

Two-Way Analysis of High-Dimensional Collinear Data^{*}

Ilkka Huopaniemi^{1,**}, Tommi Suvitaival¹, Janne Nikkilä^{1,2}, Matej Orešič³,
and Samuel Kaski¹

¹ Department of Information and Computer Science, Helsinki University of
Technology, P.O. Box 5400, FI-02015 TKK, Finland

{ilkka.huopaniemi, tommi.suvitaival, janne.nikkila, samuel.kaski}@tkk.fi

² Department of Basic Veterinary Sciences, (Division of Microbiology and
Epidemiology), Faculty of Veterinary Medicine, University of Helsinki, P.O. Box 66,
University of Helsinki FIN-00014 Finland

³ VTT Technical Research Centre of Finland, P.O. Box 1000, FIN-02044 VTT,
Espoo, Finland

matej.oresic@vtt.fi

<http://www.cis.hut.fi/projects/mi/>

Abstract. We present a Bayesian model for two-way ANOVA-type analysis of high-dimensional, small sample-size datasets with highly correlated groups of variables. Modern cellular measurement methods are a main application area; typically the task is differential analysis between diseased and healthy samples, complicated by additional covariates requiring a multi-way analysis. The main complication is the combination of high dimensionality and low sample size, which renders classical multi-variate techniques useless. We introduce a hierarchical model which does dimensionality reduction by assuming that the input variables come in similarly-behaving groups, and performs an ANOVA-type decomposition for the set of reduced-dimensional latent variables. We apply the methods to study lipidomic profiles of a recent large-cohort human diabetes study.

Keywords: ANOVA, factor analysis, hierarchical model, metabolomics, multi-way analysis, small sample-size.

Reference

1. Huopaniemi, I., Suvitaival, T., Nikkilä, J., Orešič, M., Kaski, S.: Two-Way Analysis of High-Dimensional Collinear Data. *Data Mining and Knowledge Discovery* (2009) DOI: 10.1007/s10618-009-0137-2

* This is an extended abstract of an article published in the *Data Mining and Knowledge Discovery* journal [1].

** The project was funded by Tekes MASI program. I.H., T.S and S.K belong to the Adaptive Informatics Research Centre and Helsinki Institute for Information Technology. I.H. is funded by the Graduate School of Computer Science and Engineering. S.K is partially supported by EU FP7 NoE PASCAL2, ICT 216886.