Abstract

J. Chemometrics, **30**, 153-162 (2016)
DOI: 10.1002/cem.2775

I. Hoffmann, P. Filzmoser, S. Serneels, K. Varmuza:

Sparse and robust PLS for binary classification.

Partial robust M regression (PRM), as well as its sparse counterpart sparse PRM, have been reported to be regression methods that foster a partial least squares-alike interpretation while having good robustness and efficiency properties, as well as a low computational cost. In this paper, the partial robust M discriminant analysis classifier is introduced, which consists of dimension reduction through an algorithm closely related to PRM and a consecutive robust discriminant analysis in the latent variable space.

The method is further generalized to sparse partial robust M discriminant analysis by introducing a sparsity penalty on the estimated direction vectors. Thereby, an intrinsic variable selection is achieved, which yields a better graphical interpretation of the results, as well as more precise coefficient estimates, in case the data contain uninformative variables.

Both methods are robust against leverage points within each class, as well as against *adherence outliers* (points that have been asigned a wrong class label).

A simulation study investigates the effect of outliers, wrong class labels, and uninformative variables on the proposed methods and its classical PLS counterparts and corroborates the robustness and sparsity claims. The utility of the methods is demonstrated on data from mass spectrometry analysis (time-of-flight secondary ion mass spectrometry) of meteorite samples.