

Supervised dimension reduction based on scatter matrices

Oja, Hannu* ¹

Liski, Eero ²

Nordhausen, Klaus ³

¹Department of Mathematics and Statistics, University of Turku, FI-20014 Turku,
hannu.oja@utu.fi

² Tampere School of Public Health, FI-33014, University of Tampere, eero.liski@uta.fi

³ School of Information Sciences, FI-33014, University of Tampere, klaus.nordhausen@uta.fi

Dimension reduction plays an important role in high dimensional data analysis. Principal component analysis (PCA), independent component analysis (ICA), and sliced inversion regression (SIR) are well-known but very different tools for dimension reduction. It appears that these approaches can all be seen through a comparison between two different scatter matrices S_1 and S_2 . The subspaces for dimension reduction are then given by the subspaces determined by the eigenvectors of $S_1^{-1}S_2$. Notice that in SIR the second scatter matrix is supervised in the sense that it uses the dependence between the observed random vector and a response variable. Based on these notions, we extend the invariant coordinate selection (ICS), allowing the second scatter matrix S_2 to be supervised; supervised ICS can then be used in supervised dimension reduction. It is remarkable that most supervised dimension reduction methods proposed in the literature can be reformulated in this way. In this talk, several families of supervised scatter matrices are discussed, and their use in supervised dimension reduction is illustrated with examples and simulations.

Key words: Invariant coordinate selection (ICS); Independent component analysis (ICA); Sliced inverse regression (SIR).