Contributions to dimension reduction in regression problems

Stéphane Girard

Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.

Stephane.Girard@inria.fr

Abstract: This report summarizes my contributions to regression methods in high-dimensional settings.

Standard regression methods (linear or nonlinear) suffer from the so-called curse of dimensionality when the dimension of the covariate is large (compared to the sample size). Sliced Inverse Regression (SIR) is an effective method for dimension reduction in such high-dimensional regression problems. It has been successfully applied to estimation of reference curves [1, 2].

The original method, however, requires the inversion of the predictors covariance matrix. In case of collinearity between these predictors or small sample sizes compared to the dimension, the inversion is not possible and a regularization technique has to be used. The proposed approach is based on a Fisher Lecture given by R.D. Cook where it is shown that SIR axes can be interpreted as solutions of an inverse regression problem. A Gaussian prior is introduced on the distribution on the unknown parameters of the inverse regression problem in order to regularize their estimation [3]. I showed that some existing SIR regularizations can enter this framework, which permits a global understanding of these methods [4]. Three new priors are proposed leading to new regularizations of the SIR method. A comparison on simulated data as well as an application to the estimation of Mars surface physical properties from hyperspectral images are provided [5, 6, 7].

A sequential extension of SIR dedicated to very large datasets [8] was also developped as well as an extension to multivariate outputs [9]. I also worked on the adaptation of SIR to non-Gaussian datasets [10, 11] via mixtures and heavy-tailed distributions.

References

[1] A. Gannoun, S. Girard, C. Guinot, and J. Saracco. Reference ranges based on nonparametric quantile regression. *Statistics in Medicine*, 21(20):3119–3135, 2002.

[2] A. Gannoun, S. Girard, C. Guinot, and J. Saracco. Sliced inverse regression in reference curves estimation. *Computational Statistics and Data Analysis*, 46(1):103–122, 2004.

[3] C. Bernard-Michel, L. Gardes, and S. Girard. Gaussian regularized sliced inverse regression. *Statistics and Computing*, 19:85–98, 2009.

[4] C. Bernard-Michel, L. Gardes, and S. Girard. A note on sliced inverse regression with regularizations. *Biometrics*, 64:982–986, 2008.

[5] M. Fauvel, S. Girard, S. Douté, and L. Gardes. Machine learning methods for the inversion of hyperspectral images. In A. Reimer, editor, *Horizons in World Physics*, volume 290, pages 51–77. Nova Science, New-York, 2017.

[6] S. Girard and J. Saracco. An introduction to dimension reduction in nonparametric kernel regression. In D. Fraix-Burnet and D. Valls-Gabaud, editors, *Regression methods for astrophysics*, volume 66, pages 167–196. EDP Sciences, 2014.

[7] C. Bernard-Michel, S. Douté, M. Fauvel, L. Gardes, and S. Girard. Retrieval of Mars surface physical properties from Omega hyperspectral images using regularized sliced inverse regression. *Journal of Geophysical Research - Planets*, 114, 2009. E06005.

[8] M. Chavent, S. Girard, V. Kuentz-Simonet, B. Liquet, T. M. N. Nguyen, and J. Saracco. A sliced inverse regression approach for data stream. *Computational Statistics*, 29:1129–1152, 2014.

[9] R. Coudret, S. Girard, and J. Saracco. A new sliced inverse regression method for multivariate response. *Computational Statistics and Data Analysis*, 77:285–299, 2014.

[10] A. Chiancone, S. Girard, and J. Chanussot. Collaborative sliced inverse regression. *Communication in Statistics - Theory and Methods*, 46(12):6035–6053, 2017.

[11] A. Chiancone, F. Forbes, and S. Girard. Student sliced inverse regression. *Computational Statistics and Data Analysis*, 113:441–456, 2017.