Scientific results in 2009

Stéphane Girard

Inria Grenoble Rhône-Alpes & LJK (team MISTIS).

655, avenue de l'Europe, Montbonnot. 38334 Saint-Ismier Cedex, France

Stephane.Girard@inria.fr

Abstract

This short note summarizes my scientific results in 2009. Three main research topics are addressed: High dimensional statistical learning, extreme-value analysis, and Boundary or frontier estimation.

1 High dimensional statistical learning

I have proposed a parametrization of the Gaussian mixture model for classification purposes. It is assumed that the high-dimensional data live in subspaces with intrinsic dimensions smaller than the dimension of the original space and that the data of different classes live in different subspaces with different intrinsic dimensions. New high-dimensional data classifiers [1] are introduced on the basis of this model [2, 3] in the semi-supervised context.

I also developped dimension reduction methods for high dimensional regression problems [4, 5, 6, 7] with applications to astrophysics.

Copula provides a relevant tool to build multivariate probability laws, from fixed marginal distributions and required degree of dependence. From Sklar's Theorem, the dependence properties of a continuous multivariate distribution can be entirely summarized, independently of its margins, by a copula. I have introduced a new semiparametric family of bivariate copulas. The family is generated by a univariate function, determining the symmetry (radial symmetry, joint symmetry) and dependence property (quadrant dependence, total positivity, ...) of the copulas [8].

2 Extreme-value analysis

The decay of the survival function $P(X > x)$ is driven by a real parameter called the extreme-value index. When this parameter is positive, the survival function is said to be heavy-tailed, see [9] for an application to environmental data or [10] for the modelling of network activity. When this parameter is negative, the survival function vanishes above its right end point. If

this parameter is zero, then the survival function decreases to zero at an exponential rate. An important part of my work is dedicated to the study of such distributions [11]. For instance, in reliability, the distributions of interest are included in a semi-parametric family whose tails are decreasing exponentially fast. These so-called Weibull-tail distributions include Gaussian, gamma, exponential and Weibull distributions, among others.

3 Boundary or frontier estimation

In image analysis, the boundary estimation problem arises in image segmentation as well as in supervised learning. Two different and complementary approaches are developed. In the extreme quantiles approach, approach, the boundary bounding the set of points is viewed as the larger level set of the points distribution. Its estimation is thus an extreme quantile curve estimation problem. Estimators based on projection as well as on kernel regression methods are applied on the extreme values set [12, 13].

Besides, the use of high order moments techniques permits to use all the observations from the sample [14] similarly to the methods used for the production frontier estimation in econometrics.

References

[1] C. Bouveyron and S. Girard. Classification supervisée et non supervisée des données de grande dimension. *La revue de Modulad*, 40:81–102, 2009.

[2] C. Bouveyron and S. Girard. Robust supervised classification with mixture models: Learning from data with uncertain labels. *Pattern Recognition*, 42(11):2649–2658, 2009.

[3] C. Bouveyron, S. Girard, and M. Olteanu. Supervised classification of categorical data with uncertain labels for DNA barcoding. In *17th European Symposium on Artificial Neural Networks*, pages 39–34, Bruges, Belgique, avril 2009.

[4] 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.

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

[6] C. Bernard-Michel, S. Douté, M. Fauvel, L. Gardes, and S. Girard. Machine learning techniques for the inversion of planetary hyperspectral images. In *1st IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, Grenoble, aout 2009.

[7] C. Bernard-Michel, S. Douté, M. Fauvel, L. Gardes, and S. Girard. Support vectors machines regression for estimation of Mars surface physical properties. In *17th European Symposium on Artificial Neural Networks*, pages 195–200, Bruges, Belgique, avril 2009.

[8] C. Amblard and S. Girard. A new extension of bivariate FGM copulas. *Metrika*, 70:1–17, 2009.

[9] C. Bernard-Michel, L. Gardes, S. Girard, and G. Molinié. Spatial analysis of extreme rainfalls in the Cévennes-Vivarais region. In *Spatial Extremes, Theory and Applications*, Lisbonne, Portugal, avril 2009.

[10] P. Loiseau, P. Gonçalves, S. Girard, F. Forbes, and P. Primet Vicat-Blanc. Maximum likelihood estimation of the flow size distribution tail index from sampled packet data. In *SIGMETRICS–Joint International Conference on Measurement and Modeling of Computer Systems*, Seattle, USA, juin 2009.

[11] L. Gardes, S. Girard, and A. Guillou. A unified statistical model for Pareto and Weibull tail distributions. In *6th International Conference on Extreme Value Analysis*, Fort Collins, USA, juin 2009.

[12] A. Daouia, L. Gardes, S. Girard, and A. Lekina. Extreme level curves of heavy-tailed distributions. In *6th International Conference on Extreme Value Analysis*, Fort Collins, USA, juin 2009.

[13] L. Gardes, S. Girard, and A. Lekina. Estimation non-paramétrique des quantiles extrêmes conditionnnels. In *41èmes Journées de Statistique organisées par la Société Française de Statistique*, Bordeaux, mai 2009.

[14] S. Girard and P. Jacob. Frontier estimation with local polynomials and high power-transformed data. *Journal of Multivariate Analysis*, 100:1691–1705, 2009.