My contributions to research in 2007

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 report summarizes my research results in 2007. Three main research topics are addressed: Boundary or frontier estimation, High dimensional statistical learning and Extreme-value analysis.

1 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. I focussed on the situation where a covariate is recorded simultaneously the variable of interest. In this case, the extreme-value index and the extreme quantile depend on the covariate [6]. It may be the case in hydrology for instance, see [7, 8] for an application to the study of extreme rainfalls.

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. 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. I also introduced a test on the tail behavior of a distribution [9].

2 Boundary / frontier estimation

In image analysis, the boundary estimation problem arises in image segmentation as well as in supervised learning. Here, we focus on an extreme quantiles approach where 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 kernel regression methods are applied on the extreme values set [1].

3 High dimensional statistical learning

Image analysis and computer vision are two important application domains for high dimensional data analysis and, more precisely, for dimension reduction methods. Indeed, a $M \times M$ grey-level image can be represented as a p-dimensional vector with $p = M^2$ or by a set of local descriptors. In both case, even with moderate image sizes, one obtains data living in very high-dimensional spaces. Principal Component Analysis (PCA) is usually an efficient tool for reducing the dimension of such data. However, even simple transformations between images can yield strong non-linearities in the p-dimensional space and thus strongly reduce the PCA efficiency. To overcome this problem, I have introduced Auto-Associative models allowing to build new nonlinear dimension reduction methods. The dataset is approximated by a differentiable manifold generalizing PCA's linear subspaces. The approximation algorithm is simple: it consists in incrementing the dimension of the manifold step by step [2].

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 are introduced on the basis of this model [3, 4].

I also developped dimension reduction methods for high dimensional regression problems [5].

References

[1] S. Girard and P. Jacob. Boundary estimation via regression on high power-transformed data. In *56th Session of the International Statistical Institute*, Lisbonne, aout 2007.

[2] S. Girard and S. Iovleff. Auto-associative models, nonlinear principal component analysis, manifolds and projection pursuit. In A. Gorban et al., editor, *Principal Manifolds for Data Visualisation and Dimension Reduction*, volume 28, pages 205–222. LNCSE, Springer-Verlag, 2007.

[3] C. Bouveyron, S. Girard, and C. Schmid. High dimensional data clustering. *Computational Statistics and Data Analysis*, 52:502–519, 2007.

[4] C. Bouveyron, S. Girard, and C. Schmid. High dimensional discriminant analysis. *Communication in Statistics - Theory and Methods*, 36(14):2607–2623, 2007.

[5] C. Bernard-Michel, S. Douté, L. Gardes, and S. Girard. Estimation of Mars surface physical properties from hyperspectral images using the SIR method. In *International Symposium on Applied Stochastic Models and Data Analysis*, Chania, Crete, mai 2007.

[6] L. Gardes and S. Girard. Some nonparametric estimators of the conditional tail index. In *Fifth International Symposium on Extreme Value Analysis*, Berne, juillet 2007.

[7] L. Gardes and S. Girard. Nonparametric estimation of the conditional tail index. In *Statistical Extremes and Environmental Risk Workshop*, pages 47–50, Lisbonne, février 2007.

[8] C. Conteduca, G. Molinié, S. Girard, and L. Gardes. Severe storms in mountainous Mediterranean regions: Uncertainties on extreme rainfall estimations. In *9th Plinius Conference on Mediterranean Storms*, Varenna, Italie, septembre 2007.

[9] J. Diebolt, M. Garrido, and S. Girard. A goodness-of-fit test for the distribution tail. In M. Ahsanulah and S. Kirmani, editors, *Extreme Value Distributions*, pages 95–109. Nova Science, New-York, 2007.