Research results in 2006

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 2006. Three main research topics are addressed: Boundary or frontier estimation, High dimensional statistical learning and Extreme-value analysis.

1 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 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 [1, 2].

Besides, the use of optimization techniques permits to select automatically the relevant points from all the observations of the sample [3] similarly to the methods used in Support Vector Machines (SVM).

2 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 are introduced on the basis of this model [4, 5, 6, 7].

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 [8].

Another aspect of multivariate data analysis is the modeling of dependence between variables. 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. We 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. An estimation procedure has also been introduced [9].

3   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. 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 [10, 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. I also introduced an estimator of the extreme-value index valid in all cases [12].

References

[1] J. Geffroy, S. Girard, and P. Jacob. Asymptotic normality of the L_1-error of a boundary estimator. *Nonparametric Statistics*, 18(1):21–31, 2006.

[2] S. Girard and L. Menneteau. Estimation of star-shaped supports via smoothed extreme value estimators of non-uniform point processes boundaries. In *IMS Annual Meeting and X Brazilian School of Probability*, Rio de Janeiro, Brésil, juillet 2006.

[3] S. Girard, A. Iouditski, and A. Nazin. On optimal and adaptive non-parametric estimation for periodic frontier via linear programming. In *Third International Control Conference*, Moscou, Russie, juin 2006.

[4] C. Bouveyron, S. Girard, and C. Schmid. Class-specific subspace discriminant analysis for high-dimensional data. In C. Saunder et al., editor, *Lecture Notes in Computer Science*, volume 3940, pages 139–150. Springer-Verlag, Berlin Heidelberg, 2006.

[5] C. Bouveyron, J. Kannala, C. Schmid, and S. Girard. Object localization by subspace clustering of local descriptors. In *5th Indian Conference on Computer Vision, Graphics and Image Processing*, Madurai, Inde, décembre 2006.

[6] C. Bouveyron, S. Girard, and C. Schmid. High dimensional data clustering. In *Compstat, 17th symposium of the IASC*, Rome, Italie, aout 2006.

[7] C. Bouveyron, S. Girard, and C. Schmid. Classification des données de grande dimension: application à la vision par ordinateur. In *2èmes Rencontres Inter-Associations*, pages 24–25, Lyon, mars 2006.

[8] S. Girard and S. Iovleff. Auto-associative models and generalized principal component analysis. In *Workshop principal manifolds for data cartography and dimension reduction*, Leicester, UK, aout 2006.

[9] C. Amblard and S. Girard. A semiparametric family of bivariate copulas: dependence properties and estimation procedures. In *IMS Annual Meeting and X Brazilian School of Probability*, Rio de Janeiro, Brésil, juillet 2006.

[10] L. Gardes and S. Girard. Comparison of Weibull tail-coefficient estimators. *REVSTAT - Statistical Journal*, 4(2):373–188, 2006.

[11] S. Girard. Statistical inference for Weibull tail-distributions. In *Journées Extrêmes à Lille*, Lille, mars 2006.

[12] L. Gardes and S. Girard. Asymptotic properties of a Pickands type estimator of the extreme value index. In Louis R. Velle, editor, *Focus on probability theory*, pages 133–149. Nova Science, New-York, 2006.