The effect of dependence in selecting smoothing parameters in a nonparametric regression model

Supervisors: Sana Louhichi[1] and Didier Girard

It is a Research or an Engineer project which will be held at the LJK Laboratory of the University Grenoble Alpes.

Keywords: cross-validation, time series, nonparametric regression, monte carlo studies.

Prerequisites: This project can be seen as an engineer internship or as a subject of research that may eventually lead to a thesis subject. In both cases, it is desirable that the candidate has an idea about the non-parametric statistic and that she/he is able to do simulations. The codes and the simulations will be elaborated with the software R, possibly in mathematica.

The estimators in nonparametric statistics depend on smoothing parameters. The choice of those parameters are crucial for the performance of the estimators. Many criteria for the "optimal" adaptation of linear curve -or surface- estimators have been proposed in the literature when the observations are independent. These known criteria, for selecting the smoothing parameters, used in the independence context do not work properly in the situation where the data are dependent. In fact, these criteria tend to underestimate the smoothing parameters when the observations are dependent.
The ambitious goals of this traineeship are,
(-) to study, using simulations, the ineffectiveness of the criteria introduced in the independence setting when applied to dependent data.
(-) to compare, using simulations, between different criteria introduced in the dependent case for choosing the optimal smoothing parameters.
(-) to prove some theoretical results that may explain the conclusions of these simulations.
We will focus on a nonparametric regression model:

$$Y_i = g(x_i) + \epsilon_i, \;\; 1 \leq i \leq n,$$

with an interest to,
(-) negative dependence type of the errors: ϵ is a Gaussian vector with $\mathrm{Cov}(\epsilon_i, \epsilon_j) < 0$ for any $i \neq j$.
(-) positive dependence type of the errors: ϵ is a Gaussian vector with $\mathrm{Cov}(\epsilon_i, \epsilon_j) > 0$ for any $i \neq j$.
(-) errors with Matérn-type correlations.

Some References.

1. C.-K. Chu and J. S. Marron (1991). Comparison of Two Bandwidth Selectors with Dependent Errors. Ann. Statist. Volume 19, Number 4, 1906-1918.

2. Didier A. Girard (2013). Nonparametric Curve Estimation by Kernel Smoothers: Efficiency of Unbiased Risk Estimate and GCV Selectors.
http://demonstrations.wolfram.com/NonparametricCurveEstimationByKernelSmoothersEfficiencyOfUnb/
Wolfram Demonstrations Project Published: January 9 2013.

3. Didier A. Girard (2010). Estimating the accuracy of (local) cross-validation via randomised GCV choices in kernel or smoothing spline regression. Journal of Nonparametric Statistics, American Statistical Association, 22 (1), pp.41-64.

4. Y. Wang (1998). Smoothing Spline Models with Correlated Random Errors, Journal of the American Statistical Association. Vol. 93, No. 441, pp. 341-348.

[1]E.mail: sana.louhichi@univ-grenoble-alpes.fr