Smoothing Parameter Selection When Errors are Correlated and Application to Ozone Data

by

Robert Jr St-Aubin

B.Sc, Université de Montréal, 1996

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE STUDIES

(Department of Statistics)

we accept this thesis as conforming
to the required standard

The University of British Columbia

0ctober 1998

© Robert Jr St-Aubin, 1998

Abstract

Automatic smoothing parameter selection methods for nonparametric regression like cross-validation and generalized cross-validation are known to be severely affected by dependence in the regression errors. We proposed, in this work, to modify some of the ideas used in the cross-validation criterion in kernel regression with dependent errors and apply them to smoothing splines with model based penalty. Model based penalty smoothing permits us to keep the flexibility of the nonparametric methods while it also allows us to specify a favoured parametric model which can help improve on the estimate of the regression function.

We consider the "modified cross-validation" (also known as Leave$-2l+1$ out) and the "blockwise cross-validation" smoothing parameter selection techniques which were initially proposed by Härdle and Vieu (1992) and Wehrly and Hart (1988) respectively. These two smoothing parameter selection techniques take correlation into account and alleviate its effect on the regression function estimation.

We use a simulation study to evaluate the performance of our two smoothing parameter selection techniques. We compare the results with a

few commonly used parametric techniques. Our techniques are also applied to an air pollution data set where we estimate the underlying trend of daily and monthly ground ozone levels in southern Ontario.

Contents

List of Tables

List of Figures

Acknowledgements

Foremost, I would like to thank my supervisor, Nancy Heckman, as I couldn't have chosen a better person to work with. Nancy was always willing to give advice and most importantly encouragement, all of which were much needed.

Secondly, I am grateful to Jim Zidek for approaching me with this project. I am also thankful to Jim for reading my thesis and for giving useful comments.

A huge thank also to Harry Joe for his invaluable help with C programming and with many computational aspects of my thesis.

Lastly, I would like to thank everyone in the department of Statistics at UBC for making my two years such a pleasant experience.

Robert Jr St-Aubin

The University of British Columbia

October 1998

A ma famille, à la vie qui est souvent si belle et aux S&R....