SPLINE SMOOTHING USING ROBUST GENERALIZED CROSS-VALIDATION

Mark A. Lukas[1], *Frank R. de Hoog*[2], *Robert S. Anderssen*[3]

[1] *Mathematics and Statistics, Murdoch University, South Street, Murdoch WA 6150, Australia, M.Lukas@murdoch.edu.au*

[2] *CSIRO Mathematics, Informatics and Statistics, GPO Box 664, Canberra ACT 2601, Australia, Frank.deHoog@csiro.au*

[3] *CSIRO Mathematics, Informatics and Statistics, GPO Box 664, Canberra ACT 2601, Australia, Bob.Anderssen@csiro.au*

Generalized cross-validation (GCV) is a popular criterion for the selection of the parameter λ in spline smoothing of noisy data. However, it can be unstable and sometimes leads to severe undersmoothing, especially if the sample size n is small. This shortcoming of GCV led to the development of the robust GCV (RGCV) criterion (Robinson and Moyeed 1989), which uses a combination of the GCV score function and $tr(A^2(\lambda))$, where $A(\lambda)$ is the smoothing matrix, with weighting determined by a robustness parameter $\gamma \in (0,1)$. Although RGCV was first proposed over 20 years ago, there has been little investigation of it until recently. In this talk we will discuss recent work showing that for uncorrelated data, RGCV is a practical and effective parameter selection criterion for any size n. Our development of new $O(n)$ algorithms for the calculation of $tr(A^2(\lambda))$ makes it feasible to compute the RGCV score for large n (Lukas, de Hoog and Anderssen 2010). In the analysis of RGCV (Lukas, de Hoog and Anderssen 2008), we use a geometric approach due to Efron to explain the small-sample stability of RGCV. We also derive expressions for the asymptotic inefficiency for both the prediction error and a stronger Sobolev error (involving derivatives) which show that RGCV performs well for any $\gamma \in [0.2, 0.4]$, and better than GCV. We will illustrate the results using simulations with cubic smoothing splines.

References

Lukas, M.A., de Hoog, F.R., and Anderssen, R.S. (2008). Spline smoothing using robust GCV. Report 08-154, CMIS, CSIRO.

Lukas, M.A., de Hoog, F.R., and Anderssen, R.S. (2010). Efficient algorithms for robust GCV spline smoothing. J. Comput. Appl. Math., to appear.

Robinson, T. and Moyeed, R. (1989). Making robust the cross-validatory choice of smoothing parameter in spline smoothing regression, Comm. Statist. Theory Methods, 18, 523-539.

Mark Lukas *is a Senior Lecturer in Mathematics and Statistics at Murdoch University, where he has been employed since 1985. He has taught units on calculus, discrete mathematics, and applied and computational mathematics, and has supervised several Honours and PhD projects. His research interests are regularization methods for inverse and ill-posed problems, non-parametric smoothing of noisy data and computational mathematics in general.*