Multiple predictor smoothing methods for sensitivity analysis: Example results

Curtis B. Storlie[a], Jon C. Helton[b],*

[a]Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA
[b]Department of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287-1804, USA

Abstract

The use of multiple predictor smoothing methods in sampling-based sensitivity analyses of complex models is investigated. Specifically, sensitivity analysis procedures based on smoothing methods employing the stepwise application of the following nonparametric regression techniques are described in the first part of this presentation: (i) locally weighted regression (LOESS), (ii) additive models, (iii) projection pursuit regression, and (iv) recursive partitioning regression. In this, the second and concluding part of the presentation, the indicated procedures are illustrated with both simple test problems and results from a performance assessment for a radioactive waste disposal facility (i.e., the Waste Isolation Pilot Plant). As shown by the example illustrations, the use of smoothing procedures based on nonparametric regression techniques can yield more informative sensitivity analysis results than can be obtained with more traditional sensitivity analysis procedures based on linear regression, rank regression or quadratic regression when nonlinear relationships between model inputs and model predictions are present.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Additive models; Epistemic uncertainty; Locally weighted regression; Nonparametric regression; Projection pursuit regression; Recursive partitioning regression; Scatterplot smoothing; Sensitivity analysis; Stepwise selection; Uncertainty analysis

1. Introduction

The first part of this presentation [1] reviews parametric and nonparametric regression procedures for use in sampling-based sensitivity analyses. Specifically, the following parametric regression procedures are introduced and briefly described in Section 2 of Ref. [1]: (i) linear regression (LIN_REG), (ii) rank regression (RANK_REG), and (iii) quadratic regression (QUAD_REG). Further, the following nonparametric regression procedures are introduced and briefly described in Section 3.3 of Ref. [1]: (i) locally weighted regression (LOESS), (ii) additive models (GAMs), (iii) projection pursuit regression (PP_REG), and (iv) recursive partitioning regression (RP_REG). In addition, algorithms for the stepwise implementation of these procedures in the R language as part of a sensitivity analysis are described in Section 4 of Ref. [1].

The efficacy of the various methods described in Ref. [1] as procedures for sensitivity analysis is now investigated with both analytic test model data and real data. The analytic test models were assembled as part of a review volume on sensitivity analysis [2,3]. The real data comes from a performance assessment for the Waste Isolation Pilot Plant (WIPP) [4,5]. The methods are compared on the basis of fidelity to the data, overfitting of the data, and reproducibility.

The presentation is organized as follows. First, certain quantities used in assessing the efficacy of the various sensitivity analysis procedures are introduced (Section 2). Then, the results obtained with the analytic test models are presented (Section 3). Next, the results obtained with the data from the WIPP performance assessment are presented (Section 4). The presentation then ends with a summary of observations and insights (Section 5).

*Corresponding author. Department 1544, MS 0779, Sandia National Laboratories, Albuquerque, NM 87185-0779 USA. Tel.: 505 284 4808; fax: 505 844 2348.

E-mail address: jchelto@sandia.gov (J.C. Helton).

2. Assessment of analysis efficacy

The R^2 statistic (see Eq. (2.5), Ref. [1]) provides one measure of the fidelity of a regression model to the data from which it was constructed. In particular, the closer R^2 is to one, the better the model reproduces the data. However, the R^2 statistic can be misleading in that its value can be unrealistically inflated by overfitting the data. The adjusted R^2 statistic R_A^2 provides a measure of fidelity that attempts to correct the effects of overfitting the data (pp. 91–92, Ref. [6]). Specifically, R_A^2 is defined by

$$R_A^2 = 1 - \frac{(nS-1)\sum_{i=1}^{nS}(\hat{y}_i - y_i)^2}{(nS-p)\sum_{i=1}^{nS}(y_i - \bar{y})^2} = 1 - (1-R^2)\left(\frac{nS-1}{nS-p}\right),$$

(2.1)

where p is the number of degrees of freedom associated with the fitted model. However, the values for R^2 and R_A^2 are similar when p is small relative to nS.

The PRESS statistic PRS (see Eq. (3.19), Ref. [1]) provides a way to test for an overfitting of the data. In particular, a decrease in PRS with the addition of a variable to a model indicates an improvement in the predictive capability of the model (i.e., the fidelity of model to the data has increased). In contrast, an increase in PRS indicates that an overfitting of the data has taken place. This property results because the PRESS statistic is very sensitive to the effects of a limited number of highly influential observations (typically observations with extreme values for one or a few independent variables). Monitoring PRESS values as variables are added to a model provides a way to check for overfitting of the data, with such overfitting indicated when the addition of a variable results in an increase in the PRESS value over the PRESS value obtained before the addition of that variable. Such a jump in the PRESS statistic indicates the model is starting to "chase" results associated with individual observations rather than following actual patterns in the data.

The PRESS statistic can also be used to compare the fidelity of models constructed from the same data set but with different procedures. In particular, a model with a lower PRESS value is preferable to a model with a higher PRESS value. However, there are two drawbacks in using PRESS to compare models obtained with different procedures. First, PRESS values can be very sensitive to the effects of a limited number of extreme observations. Second, there is no "absolute" standard against which a PRESS value can be compared to indicate whether or not a model is providing a good match to the data. In contrast, R^2 values approach one as the fidelity of the model to the data increases; unfortunately, there is no such limiting value for the PRESS statistic.

The adjusted PRESS value PRS_A (see Eqs. (3.22)–(3.25), Ref. [1]) reduces the effects of highly influential observations by using an average leverage value in its definition. The adjusted PRESS value PRS_A is similar in concept to

the adjusted R^2 value R_A^2 in that it penalizes a model for the use of an excessive number of degrees of freedom in its construction. However, as with the original statistic PRS, there is no limiting value for PRS_A that provides a standard by which the fidelity of a model to the underlying data can be judged. Although PRS_A can be more useful than the original PRESS statistic in comparing models constructed with different procedures, it is less effective in checking for overfitting because of the reduction in the effects of extreme observations.

The top-down coefficient of concordance (TDCC) provides a way to assess the reproducibility of sensitivity analysis results obtained with individual analysis procedures [7,8]. In particular, the TDCC provides a measure of the agreement between results obtained with independently generated samples in a manner that emphasizes agreement in the identification of the most important variables and places less emphasis on agreement in the identification of the less important variables. For notational purposes in the definition of the TDCC, suppose (i) nR independently generated samples of the same size involving a vector $\mathbf{x} = [x_1, x_2, \ldots, x_{nX}]$ of independent variables are under consideration, (ii) a sensitivity analysis to rank variable importance is carried out for each sample, and (iii) r_{jk} denotes the rank assigned to variable j in the indicated sensitivity analysis for sample k, where the most important variable is assigned a rank of 1, the next most important variable is assigned a rank of 2, and so on, with variables of the same importance assigned their average rank (the preceding is the reverse of the ranking procedure described in Section 2.2 of Ref. [1] for rank regression). The TDCC is then defined by

$$C_T$$
$$= \left\{\sum_{j=1}^{nX}\left[\sum_{k=1}^{nR} ss(r_{jk})\right]^2 - (nR)^2 nX\right\}\Bigg/\left\{(nR)^2\left(nX - \sum_{j=1}^{nX} 1/j\right)\right\},$$

(2.2)

where $ss(r_{jk})$ is the Savage score given by

$$ss(r_{jk}) = \sum_{i=r_{jk}}^{nX} 1/i$$

for variable j in a sample k and average Savage scores are assigned in the event of ties. Use of the Savage scores $ss(r_{jk})$ rather than the ranks r_{jk} in the definition of the TDCC in Eq. (2.2) results in the previously indicated emphasis on agreement on the most important variables and deemphasis on disagreement on the less important variables.

In the examples that follow, variable importance is defined by the order in which variables enter the model under construction, with the first variable entering the model ranked 1, the second variable entering the model ranked 2, and so on. The variables that are not selected for entry into the model are all assigned the same average rank. The preceding ranking is used in the calculation of the TDCC. Values for the TDCC close to one indicate a high