Journal of Korean
Data & Information Science Society
2007, Vol. 18, No. 1, pp. 229 ~ 235

Leverage Measures in Nonlinear Regression[1]

Myung-Wook Kahng[2]

Abstract

Measures of leverage in nonlinear regression models are discussed by extending the leverage in linear regression models. The connection between measures of leverage and nonlinearity of the models are explored. Illustrative example based on real data is presented.

Keywords : Jacobian Matrix, Leverage, Perturbation, Tangent Plane Approximation

1. Introduction

It is well known that statistical inferences can be substantially influenced by one observation or a few observations in regression models; that is, not all observations have equal importance. In assessing the importance of individual observations, one is typically interested in identifying observations that have a greater-than-average impact on the estimation of model parameters and fitted values. These influential points are often identified through the use of case deletion and perturbation diagnostics.

Leverage is one of the basic components of influence in linear regression models (Chatterjee and Hadi, 1986). By Belsley, Kuh, and Welsch(1980) and Ross(1987), leverage has been generalized via linear approximation to more complex response models. Emerson, Hoaglin, and Kempthorne(1984) considered definitions of leverage for the nonlinear regression models.

Here, we discuss measures of leverage in nonlinear regression models. In Section 2, we briefly review the leverage in linear regression models. We develop the measures of leverage in nonlinear models in Section 3. In Section 4, we provide examples, and in Section 5, we discuss the relationship between these

1) This research was supported by the Sookmyung Women's University Research Grants 2005.
2) Professor, Department of Statistics, Sookmyung Women's University, Seoul, 140-742, Korea
E-mail : mwkahng@sookmyung.ac.kr

measures and nonlinearity.

2. Leverage in Linear Regression

The usual linear model can be written in matrix form as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{y}$ is $n \times 1$ vector of responses, $\boldsymbol{X}$ is the $n \times p$ matrix of fixed predictor values, $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression coefficients, and $\boldsymbol{\epsilon}$ is the vector of independent errors each with mean 0 and variance σ^2.

Assuming $\boldsymbol{X}$ has full column rank, we define the $n \times n$ hat matrix as

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$$

with elements $\boldsymbol{H} = \{h_{ij}\}$. The matrix is a linear operator which projects onto $S(\boldsymbol{X}) \subset \boldsymbol{R}^n$. It was named the "hat" matrix by John Tukey because it puts the "hat" on the vector of fitted values

$$\hat{\boldsymbol{y}} = \boldsymbol{Hy} .$$

The diagonal elements h_{ii} of the hat matrix are called leverages and generally show how sensitive the $\hat{y}_i$ is to perturbations in the corresponding observed response, y_i. Suppose we modify $\boldsymbol{y}$ by adding $b\boldsymbol{c}_i$ where b is a fixed constant and $\boldsymbol{c}_i$ indicates the i-th standard basis vector for $\boldsymbol{R}^n$ with a 1 in the i-th position and 0's elsewhere. Then the fitted values based on $\boldsymbol{y} + b\boldsymbol{c}_i$ are

$$\hat{\boldsymbol{y}}(b) = \boldsymbol{Hy} + b\boldsymbol{Hc}_i$$

so that

$$\hat{\boldsymbol{y}}(b) - \hat{\boldsymbol{y}} = b\boldsymbol{Hc}_i$$

and thus

$$\frac{\hat{\boldsymbol{y}}(b) - \hat{\boldsymbol{y}}}{b} = \boldsymbol{h}_i$$

where $\boldsymbol{h}_i$ is the i-th column of $\boldsymbol{H}$. We can therefore think of h_{ij} as the rate of change in $\hat{y}_j$ when perturbating y_i. In particular, h_{ii} is the rate of change in $\hat{y}_i$ when perturbating y_i.

There are several useful implications of these results: $0 \leqq h_{ii} \leqq 1$, so $\hat{y}_i$ can never change faster than y_i. If $h_{ii} = 1$ then $\hat{y}_i = y_i$, consequently, one parameter estimated is being determined by one observation. Increasing y_i never decreases $\hat{y}_i$. Fitted values for cases with h_{ii} relatively large are relatively more variable, $Var(\hat{y}_i) = h_{ii}\sigma^2$.

In the next section the idea of leverage in nonlinear regression is developed in a manner similar to that for linear regression.

3. Leverages in Nonlinear Regression

The standard nonlinear regression model can be expressed as

$$y_i = f(\boldsymbol{x}_i, \boldsymbol{\theta}) + \epsilon_i, \quad i = 1, \cdots, n$$

in which the i-th response y_i is related to the q-dimensional vector of known explanatory variables $\boldsymbol{x}_i$ through the known model function f, which depends on the p-dimensional unknown parameter $\boldsymbol{\theta} \in \Theta$, and ϵ_i is error. We assume that f is twice continuously differentiable in $\boldsymbol{\theta}$, and errors ϵ_i are *i.i.d* normal random variables with mean 0 and variance σ^2. In matrix notation we may write,

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\epsilon}$$

where $\boldsymbol{y}$ is an n-dimensional vector with elements $y_1, \cdots, y_n$, $\boldsymbol{X}$ is an $n \times q$ matrix with rows $x_1^T, \cdots, x_n^T$, $\boldsymbol{\epsilon}$ is an n-dimensional vector with elements $\epsilon_1, \cdots, \epsilon_n$, and $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) = (f(\boldsymbol{x}_1, \boldsymbol{\theta}), \cdots, f(\boldsymbol{x}_n, \boldsymbol{\theta}))^T = (f_1(\boldsymbol{\theta}), \cdots, f_n(\boldsymbol{\theta}))^T = \boldsymbol{f}(\boldsymbol{\theta})$. Given the response vector $\boldsymbol{y}$, the least squares estimate of $\boldsymbol{\theta}$ is denoted $\hat{\boldsymbol{\theta}}$, and the predicted response vector is $\hat{\boldsymbol{y}} = \boldsymbol{f}(\boldsymbol{X}, \hat{\boldsymbol{\theta}}) = \boldsymbol{f}(\hat{\boldsymbol{\theta}})$.

A tangent plane approximation to the expectation surface $\{\boldsymbol{f}(\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta\}$ at $\hat{\boldsymbol{\theta}}$ is used to make inferences about $\boldsymbol{\theta}$ through the derived linear model

$$\boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{f}(\hat{\boldsymbol{\theta}}) + \hat{\boldsymbol{V}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

where $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\theta}) = \partial \boldsymbol{f} / \partial \boldsymbol{\theta}^T$ is the $n \times p$ matrix and $\hat{\boldsymbol{V}} = \boldsymbol{V}(\hat{\boldsymbol{\theta}})$. Based on the this approximation, the tangent plane leverage matrix, denoted by $\hat{\boldsymbol{H}}$, can be written as

$$\hat{\boldsymbol{H}} = \hat{\boldsymbol{V}}(\hat{\boldsymbol{V}}^T \hat{\boldsymbol{V}})^{-1} \hat{\boldsymbol{V}}^T.$$

The diagonal elements $\hat{h}_{ii}$ of $\hat{\boldsymbol{H}}$ are frequently used as a measure of leverage in nonlinear regression (Ross, 1987).

Emerson, Hoaglin, and Kempthorne(1984) generalized the measures of leverage due to perturbation of response as follows. Modifying the response vector $\boldsymbol{y}$ by adding $b\boldsymbol{c}_i$ gives the perturbed response vector $\boldsymbol{y}+b\boldsymbol{c}_i$. The least square estimate of $\boldsymbol{\theta}$ for the perturbed data is denoted $\hat{\boldsymbol{\theta}}(b)$, and the perturbed predicted response vector is

$$\hat{\boldsymbol{y}}(b)=\boldsymbol{f}(\boldsymbol{X},\hat{\boldsymbol{\theta}}(b))=\boldsymbol{f}(\hat{\boldsymbol{\theta}}(b)).$$

Then, following the discussion of linear regression in the last section, we can base leverages in nonlinear regression on limits of the form

$$\lim_{b\to 0}\frac{\hat{\boldsymbol{y}}(b)-\hat{\boldsymbol{y}}}{b}.$$

To find an informative expression that allows relatively straight-forward calculation, let $\hat{\boldsymbol{\theta}}(b)$ indicate the ordinary least squares estimate of $\boldsymbol{\theta}$ with response $\boldsymbol{y}+b\boldsymbol{c}_i$, $\dot{\boldsymbol{\theta}}(0)$ and $\ddot{\boldsymbol{\theta}}(0)$ be $\partial\hat{\boldsymbol{\theta}}(b)/\partial b$ and $\partial^2\hat{\boldsymbol{\theta}}(b)/\partial b^2$ both evaluated at $b=0$, respectively. Then, expanding around $b=0$ up to terms of order two, we have

$$\begin{aligned}\hat{\boldsymbol{y}}(b)&=\boldsymbol{f}(\hat{\boldsymbol{\theta}}(b))\\&=\boldsymbol{f}(\hat{\boldsymbol{\theta}}(0))+\hat{\boldsymbol{V}}\dot{\boldsymbol{\theta}}(0)b+\frac{1}{2}b^2\left(\dot{\boldsymbol{\theta}}(0)^T\hat{\boldsymbol{W}}\dot{\boldsymbol{\theta}}(0)+\hat{\boldsymbol{V}}\ddot{\boldsymbol{\theta}}(0)\right)\end{aligned}$$

where $\boldsymbol{W}=\boldsymbol{W}(\boldsymbol{\theta})=\partial^2\boldsymbol{f}/\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T$ is the $n\times p\times p$ array and $\hat{\boldsymbol{W}}=\boldsymbol{W}(\hat{\boldsymbol{\theta}})$. Clearly,

$$\lim_{b\to 0}\frac{\hat{\boldsymbol{y}}(b)-\hat{\boldsymbol{y}}}{b}=\hat{\boldsymbol{V}}\dot{\boldsymbol{\theta}}(0).$$

The next step is to find a useful expression for $\dot{\boldsymbol{\theta}}(0)$ by using the normal estimating equations. In particular, for all b and $j=1,\cdots,p$

$$\sum_{i=1}^{n}\left(y_i+b\gamma_i-f_i(\boldsymbol{\theta})\right)\frac{\partial f_i(\boldsymbol{\theta})}{\partial\theta_j}\bigg|_{\hat{\theta}(b)}=0$$

where γ_i is the i-th element of $\boldsymbol{c}_i$. Thus, we can write

$$\sum_{i=1}^{n}\left(y_i+b\gamma_i-f_i(\hat{\boldsymbol{\theta}}(b))\right)\frac{\partial f_i(\hat{\boldsymbol{\theta}}(b))}{\partial\theta_j}=0.$$

Let $\boldsymbol{W}_i = \boldsymbol{W}_i(\boldsymbol{\theta}) = \partial^2 f_i / \partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T$ be the $p \times p$ Hessian matrix and $\widehat{\boldsymbol{W}}_i = \boldsymbol{W}_i(\hat{\boldsymbol{\theta}})$. Taking the derivative of both sides with respect to b we have

$$\begin{aligned}
&\sum_{i=1}^{n}\left(y_i + b\gamma_i - f_i(\hat{\boldsymbol{\theta}}(b))\right)\sum_{m=1}^{p}\frac{\partial^2 f_i(\hat{\boldsymbol{\theta}}(b))}{\partial\theta_j\partial\theta_m}\dot{\theta}_m(b) \\
&\qquad + \sum_{i=1}^{n}\frac{\partial f_i(\hat{\boldsymbol{\theta}}(b))}{\partial\theta_j}\left(\gamma_i - \sum_{m=1}^{p}\frac{\partial f_i(\hat{\boldsymbol{\theta}}(b))}{\partial\theta_m}\dot{\theta}_m(b)\right) \\
&= \sum_{i=1}^{n} e_i \widehat{\boldsymbol{W}}_i \dot{\boldsymbol{\theta}}(0) + \widehat{\boldsymbol{V}}^T \boldsymbol{c}_i - \widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}}\dot{\boldsymbol{\theta}}(0) \\
&= 0\ .
\end{aligned}$$

Next, solving for $\dot{\boldsymbol{\theta}}(0)$ gives

$$\dot{\boldsymbol{\theta}}(0) = (\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}} - \sum_{i=1}^{n} e_i \widehat{\boldsymbol{W}}_i)^{-1}\widehat{\boldsymbol{V}}^T\boldsymbol{c}_i$$

where $e_i = y_i - f_i(\hat{\boldsymbol{\theta}}) = y_i - \hat{y}_i$ is residual and therefore

$$\begin{aligned}
\lim_{b\to 0}\frac{\hat{\boldsymbol{y}}(b) - \hat{\boldsymbol{y}}}{b} &= \widehat{\boldsymbol{V}}(\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}} - \sum_{i=1}^{n} e_i \widehat{\boldsymbol{W}}_i)^{-1}\widehat{\boldsymbol{V}}^T\boldsymbol{c}_i \\
&= \hat{\boldsymbol{J}}\boldsymbol{c}_i
\end{aligned}$$

where

$$\hat{\boldsymbol{J}} = \widehat{\boldsymbol{V}}(\widehat{\boldsymbol{V}}^T\widehat{\boldsymbol{V}} - \sum_{i=1}^{n} e_i \widehat{\boldsymbol{W}}_i)^{-1}\widehat{\boldsymbol{V}}^T,$$

which is called the Jacobian leverage matrix.

4. Examples

The data on the metabolism of tetracycline were presented in Bates and Watts 1988) and are reproduced in Table 1. The proposed model is the following:

$$f(x,\boldsymbol{\theta}) = \theta_3\left[\exp(-\theta_1(x-\theta_4)) - \exp(-\theta_2(x-\theta_4))\right].$$

The least squares estimates of the parameters are $\widehat{\theta_1} = 0.149$, $\widehat{\theta_2} = 0.716$, $\widehat{\theta_3} = 2.650$,

$\hat{\theta}_4 = 0.412$ and $\hat{\sigma} = 0.0448$. The diagonal elements $\hat{J}_{ii}$ and $\hat{h}_{ii}$ of Jacobian leverage matrix $\hat{\boldsymbol{J}}$ and tangent plane leverage matrix $\hat{\boldsymbol{H}}$, respectively, are given in Table 1. In some cases, there is a striking disagreement between the two measures.

<Table 1> Tetracycline data and leverage measures

x_i	y_i	$\hat{J}_{ii}$	$\hat{h}_{ii}$
1	0.7	0.960	0.978
2	1.2	0.548	0.617
3	1.4	0.375	0.383
4	1.4	0.353	0.359
6	1.1	0.375	0.421
8	0.8	0.258	0.271
10	0.6	0.264	0.258
12	0.5	0.334	0.334
16	0.3	0.360	0.379

5. Remarks

The Jacobian leverage matrix differ from the more standard tangent plane leverage matrix $\hat{\boldsymbol{H}}$, which is observed directly by replacing $\boldsymbol{X}$ in linear models with $\hat{\boldsymbol{V}}$. Equivalently, the tangent plane leverage matrix follows immediately from the working linear model.

Taking $\hat{\boldsymbol{J}}$ as our definition of leverage for a nonlinear regression model, we may consider $\hat{\boldsymbol{H}}$ as an approximation to it. The appropriateness of this approximation will depend upon the adequacy of the tangent plane approximation. The difference between $\hat{\boldsymbol{J}}$ and $\hat{\boldsymbol{H}}$ may be thought of as a difference in the matrices used to form the inner product between columns of matrix $\hat{\boldsymbol{V}}$. These matrices, $(\hat{V}^T\hat{V})^{-1}$ and $(\hat{V}^T\hat{V} - \sum_{i=1}^{n} e_i W_i)^{-1}$, are proportional to the inverse of the expected information and the observed information (Kahng, 1995).

The Jacobian leverage matrix is interpreted in the same way as it is in linear regression. In particular, the i-th diagonal element $\hat{J}_{ii}$ of $\hat{\boldsymbol{J}}$ is non-zero but it may be larger than one. Cases with $\hat{J}_{ii} > 1$ are called superleverage cases because the fitted value $\hat{y}_i$ changes faster than the observed value y_i (St. Laurent and Cook, 1992).

References

1. Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*, John Wiley and Sons: New York.
2. Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity,* John Wiley and Sons, New York.
3. Chatterjee, S. and Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression, *Statistical Science,* 1, 379-416.
4. Emerson, J. D., Hoaglin, D. C., and Kempthorne, P. J. (1984). Leverage in least squares additive-plus-multiplicative fits for two-way tables, *Journal of the American Statistical Association,* 79, 329-335.
5. Kahng, M. (1995). Testing outliers in nonlinear regression, *Journal of the Korean Statistical Society,* 24, 419-437.
6. Ross, W. H. (1987). The geometry of case deletion and the assessments of inflence in nonlinear regression, *The Canadian Journal of Statistics,* 15, 91-103.
7. St. Laurent, R. T. (1992). Leverage and superleverage in nonlinear regression, *Journal of the American Statistical Association,* 87, 985-990.

[received date : Dec. 2006, accepted date : Jan. 2007]