Bi-Responses Nonparametric Regression Model

Using MARS and Its Properties

Ayub Parlin Ampulembang
Department of Statistics
Faculty of Mathematics and Natural Sciences
Institut Teknologi Sepuluh Nopember
Jalan Arif Rahman Hakim, Surabaya 60111, Indonesia

Bambang Widjanarko Otok and Agnes Tuti Rumiati
Department of Statistics
Faculty of Mathematics and Natural Sciences
Institut Teknologi Sepuluh Nopember
Jalan Arif Rahman Hakim, Surabaya 60111, Indonesia

Budiasih
Department of Applied Statistics
Sekolah Tinggi Ilmu Statistik
Jalan Otista 64C, Jakarta 13330, Indonesia

Abstract

In many cases of regression analysis, we can find unknown relationship pattern between two response variables with many predictor variables (multi-predictor) and both of responses are correlated each other. Consequently such regression problem should be solved using bi-responses nonparametric regression model, which able to overcome high dimensionality case in multi-predictor data. One of the methods is Multivariate Adaptive Regression Spline (MARS). This paper aims to study how MARS method could be utilized on bi-responses nonparametric regression model. The study begins with obtaining model equation form. Thus we estimating regression function and weighted from its model equation. Last step is
done by investigating the estimator properties of its model. Results show that the equation form of its model is \( y = B_0 + \varepsilon \), with regression function and weighted of its equation estimation are \( \hat{f}(x) = B (B^T W B)^{-1} B^T W y \) and \( \hat{W} \) matrix respectively, where the diagonal elements of \( \hat{W} \) as \( (\hat{y}_1^T A_{11} \hat{y}_1 / \hat{y}_2^T A_{22} \hat{y}_2) \) and \( (\hat{y}_2^T A_{22} \hat{y}_2 / \hat{y}_1^T A_{11} \hat{y}_1) \), whereas the off-diagonal elements of \( \hat{W} \) matrix as \( \hat{y}_1^T A_{11} \hat{y}_2 / (\hat{y}_1^T A_{12} \hat{y}_2) \) and \( \hat{y}_2^T A_{22} \hat{y}_1 / (\hat{y}_2^T A_{21} \hat{y}_1) \). The results also show that the properties of this model is a bias estimator and linear in observation \( \hat{y}_2 \).

**Keywords**: Nonparametric Regression, Bi-responses, MARS, Weighted

### 1. Introduction

Generally, regression analysis is a statistical method which could explain relationship between response variable and one or more predictor variables through regression curve. When the shape of its curve is known, regression parameters could be obtained by using parametric approach. But often the shape of relationship pattern between response and predictor variables is unknown or hard to interpret. If such case occur, nonparametic regression approach could be utilized for approaching regression curve [1]. Recently, it seems that nonparametric regression receive better attention from many researchers because it possesses higher flexibility than the parametric regression does [2]. Among many researcher which develop nonparametric regression method with single response including Lee [3] for spline function; Lin, Li and Chen [4] for polynomial local function; Hallin, Lu and Tran [5] for local linear function; Friedman [6], Otok [7], Martinez, Shih, Chen and Kim [8] for MARS function. In many cases, the regression analysis requires more than a single response. Sometimes we can also found that the relationship pattern between two response variables were correlated each other. As a result, regression problems must be solved during bi-responses regression model. There are some researchers who have studied it, using some nonparametric regression function, including Wang, Guo and Brown [9], and Fernandes, Budiantara, Otok and Suhartono [10] for spline function; Welsh and Yee [11] for local linear function and Chamidah, Budiantara, Sunaryo and Zain [12] for local polynomial function. Nevertheless, the researchers mentioned above were still reviewing bi-responses nonparametric regression models in low dimensionality case. So that arise some question of how to develop the model in high dimensionality case.

According to Hastie, Tibshirani and Friedman [13], nonparametric regression problem which is often to be encountered on using multi-predictor data is high dimensionality case (curse of dimensionality). One method which was able to
Bi-responses nonparametric regression model

overcome the case is MARS. This method was first introduced by Friedman [6] using single response, which is a complex combination between spline and recursive partitioning regression (RPR). This paper aims to study bi-responses nonparametric regression model using MARS. The study begins with obtaining the equation form of its model. The next step is to obtain estimation of regression function and weighted from its model using maximum likelihood estimation (MLE) method. Finally, the study will investigate the estimator properties of its model.

2. Bi-responses Nonparametric Regression Model

Bi-responses nonparametric regression model can be written in regression equation form as follows [9]:

\[ y_i = f_i(x_i) + \varepsilon_i, \quad \ell = 1, 2; \quad i = 1, 2, \ldots, n; \]  

where \( y \) is response variable, \( x \) is predictor variable, \( f \) is regression curve and \( \varepsilon \) is random error. Based on \( i \)-th observation for \( \ell \) response, Eq. (1) can be described in matrix form as follows:

\[ \tilde{y} = f(\tilde{x}) + \varepsilon \]

where:

\[ \tilde{y} = (y_1, y_2)^T, \quad \tilde{x} = (x_1, x_2, \ldots, x_n)^T, \quad f(\tilde{x}) = \left[ f_1(x_1), f_2(x_2), \ldots, f_{11}(x_{11}), f_{12}(x_{12}), \ldots, f_{2n}(x_{2n}) \right]^T, \]

\[ \varepsilon = (\varepsilon_1, \varepsilon_2)^T = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in})^T \]

According to Wang et al. [9], random error \( \varepsilon \) in Eq. (2) is assumed normally distributed with zero mean and variance-covariance matrix \( \theta W \) as follows:

\[ \theta W = \Sigma \otimes I_n \]

where:

\[ W = \begin{pmatrix} n I_n & \rho I_n \\ \rho I_n & I_n / r \end{pmatrix} \]

is a \( 2n \times 2n \) matrix, \( \theta = \sigma_1 \sigma_2 \) and \( r = (\sigma_1 / \sigma_2) \) are scalars,

\[ \Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \]

is a \( 2 \times 2 \) matrix and \( I_n \) is a \( n \times n \) identity matrix.

hence:

\[ \tilde{y} - N(f(\tilde{x}), \theta W) \]

The value of \( W \) in Eq. (3) serves to accommodate the presence of correlation between both responses variables.

3. MARS with Single Response

MARS method is a multivariate nonparametric regression approach which developed by Friedman in 1991. This method has several advantages including flexible for high dimensionality problems [13]. According to Friedman [6], MARS with single response can be expressed in the following equation:

\[ y_i = f(x_{i1}, x_{i2}, \ldots, x_{im}) + \varepsilon_i, \quad i = 1, 2, \ldots, n \]
\[ f = \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{i=1}^{K_m} \left( s_{im} \left( x_{j(i,m)} - t_{j(i,m)} \right) \right) + \epsilon, \quad j=1,2,\ldots,p \]  

with:  
\[ (x_{j(i,m)} - t_{j(i,m)}) \begin{cases} \left( x_{j(i,m)} - t_{j(i,m)} \right), & x_{j(i,m)} \geq t_{j(i,m)} \\ 0, & \text{other} \end{cases} \]

\( f \) is MARS function, \( \alpha \) is coefficient of basis functions, \( m = 1,2,\ldots,M \) is the \( m \)-th basis function, \( s_{km} = \pm 1 \) is a pair of basis function, \( k = 1,2,\ldots,K_m \) is the \( k \)-th interaction and \( t \) is knot. Based on \( i \)-th observation, Eq. (5) can be described in matrix form as follows:

\[ \mathbf{y} = \mathbf{B} \mathbf{\alpha} + \mathbf{\epsilon} \]

where:
\[ \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{\alpha} = \begin{pmatrix} \alpha_0, \alpha_1, \ldots, \alpha_M \end{pmatrix}, \quad \mathbf{\epsilon} = \begin{pmatrix} \epsilon_1, \epsilon_2, \ldots, \epsilon_n \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 \prod_{i=1}^{K_1} \left( s_{1i} \left( x_{j(1,1)} - t_{j(1,1)} \right) \right) & \cdots & \prod_{i=1}^{K_M} \left( s_{Mi} \left( x_{j(i,M)} - t_{j(i,M)} \right) \right) \end{pmatrix} \]

\[ 1 = (1,1,\ldots,1)^T; \quad x_j = \begin{pmatrix} x_{j1}, x_{j2}, \ldots, x_{jm} \end{pmatrix}^T; \quad j = 1,\ldots,p \]

Random error \( \mathbf{\epsilon} \) in Eq. (6) is assumed normally distributed with zero mean and variance-covariance matrix \( \sigma^2 \mathbf{I} \). Based on these assumptions, then estimation of \( \mathbf{\hat{y}} \) is obtained by least squares (LS) estimation, thus obtained:

\[ \mathbf{\hat{y}} = \mathbf{B}^T \mathbf{\hat{\alpha}} = \mathbf{B}^T \left( \mathbf{B}^T \mathbf{B} \right)^{-1} \mathbf{B}^T \mathbf{\hat{y}} = \mathbf{H}^T \mathbf{\hat{y}} \]

where \( \mathbf{H} = \mathbf{B}^T \left( \mathbf{B}^T \mathbf{B} \right)^{-1} \mathbf{B}^T \) is Hat matrix that has very important role in statistical inference nonparametric regression with MARS [7].

4. Study of Bi-responses Nonparametric Regression Using MARS

Bi-responses nonparametric regression based MARS function can be obtained by combining equation (2) with equation (6), in order to obtain a new equation as follows:

\[ \mathbf{\hat{y}} = \mathbf{\hat{f}}(\mathbf{x}) + \mathbf{\epsilon} = \mathbf{B} \mathbf{\hat{\alpha}} + \mathbf{\epsilon} \]

where:
\[ \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}^T, \quad \mathbf{x} = \begin{pmatrix} x_1, x_2, \ldots, x_p \end{pmatrix}^T = \begin{pmatrix} (x_{11}, \ldots, x_{1p}) \right( (x_{12}, \ldots, x_{p2}) \right) \ldots (x_{1n}, \ldots, x_{pn}) \right) \]
\[ \mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{pmatrix}^T \]
\[ f_1(\mathbf{x}) = \begin{pmatrix} f_{11}(x_{11}, \ldots, x_{1p}) \\ f_{12}(x_{12}, \ldots, x_{p2}) \\ \vdots \\ f_{1n}(x_{1n}, \ldots, x_{pn}) \end{pmatrix}^T \]
\[ f_2(\mathbf{x}) = \begin{pmatrix} f_{21}(x_{11}, \ldots, x_{1p}) \\ f_{22}(x_{12}, \ldots, x_{p2}) \\ \vdots \\ f_{2n}(x_{1n}, \ldots, x_{pn}) \end{pmatrix}^T \]
\[ \mathbf{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}^T = \begin{pmatrix} \epsilon_{11}, \epsilon_{12}, \ldots, \epsilon_{1n}, \epsilon_{21}, \epsilon_{22}, \ldots, \epsilon_{2n} \end{pmatrix}^T \]
\[ \mathbf{B} = \text{diag}(\mathbf{B}_1, \mathbf{B}_2) \]
Bi-responses nonparametric regression model

\[ B_1 = \left( \prod_{i=1}^{K_1} \left[ (x_{j(i),m})^t - t_{j(i),m} \right] \right) \cdots \left( \prod_{i=1}^{K_2} \left[ (x_{j(i),m}^2) - t_{j(i),m}^2 \right] \right) \]

\[ B_2 = \left( \prod_{i=1}^{K_1} \left[ (x_{j(i),m})^t - t_{j(i),m} \right] \right) \cdots \left( \prod_{i=1}^{K_2} \left[ (x_{j(i),m}^2) - t_{j(i),m}^2 \right] \right) \]

\[ 1 = (1, 1, \ldots, 1) ; \quad x_j = (x_{j1}, x_{j2}, \ldots, x_{jp})^T ; \quad j = 1, 2, \ldots, p \]

\[ \alpha = (\alpha_1, \alpha_2)^T = \left( \alpha_{10}, \alpha_{11}, \alpha_{12}, \ldots, \alpha_{1M}, \alpha_{20}, \alpha_{21}, \alpha_{22}, \ldots, \alpha_{2M} \right)^T \]

Following the assumption in Eq. (4), then \( \varepsilon \) in Eq. (8) is assumed to be normally distributed with zero mean and variance-covariance matrix \( \theta W \), so that the assumption of \( \varepsilon \) can be written as:

\[ y \sim N(f(x), \theta W) \]  \hspace{1cm} (9)

based on \( y_1 \) and \( y_2 \). Eq. (9) can be rearranged as:

\[ \left( \begin{array}{c} y_1 \\ y_2 \end{array} \right) \sim N \left( \begin{array}{c} f_1(x) \\ f_2(x) \end{array} \right) \Sigma \otimes I_s \]  \hspace{1cm} (10)

where \( f_1(x) = B_1 \alpha_1 \) and \( f_2(x) = B_2 \alpha_2 \). Based on those assumptions, then estimation of \( f(x) \) in Eq. (8), can be obtained by MLE method given by following theorem 1.

**Theorem 1.**

If the bi-responses nonparametric regression model based on MARS function is expressed in Eq. (8) with its assumption in Eq. (9), then estimation of its regression function is:

\[ \hat{f}(x) = H_y \]  \hspace{1cm} (11)

where \( H = B (B^TW^{-1}B)^{-1}B^TW^{-1} \).

**Proof:**

Based on assumption in Eq. (9), then the likelihood function of \( y \) in Eq. (8), can be written as follows:

\[ L(B\alpha, \theta W) = (2\pi)^{-n}(\theta W)^{-\frac{1}{2}}e^{tr \left( -\frac{1}{2\theta}W^{-1}(y-B\alpha) (y-B\alpha)^T \right) } \]

and the log of the likelihood function is:

\[ \ell_n(B\alpha, \theta W) = -n \log(2\pi) - \frac{1}{2} \log(\theta W) - \frac{1}{2\theta}tr \left( W^{-1}(y-B\alpha) (y-B\alpha)^T \right) \]

Parameter estimation of \( \alpha \) can be obtained by setting the partial derivative of \( \ell_n(B\alpha, \theta W) \) with respect to \( \alpha \) equal to zero, as follows:

\[ \frac{\partial}{\partial \alpha} \ell_n(B\alpha, \theta W) = -\frac{1}{2\theta} \left( W^{-1}(y-B\alpha) (y-B\alpha)^T \right) = 0 \]  \hspace{1cm} (12)
By using theorem in Gupta and Nagar [14], we obtain:

\[ tr \left[ W^{-1} \left( \mathbf{y} - B\alpha \right) \left( \mathbf{y} - B\alpha \right)^T \right] = tr \left[ \left( \mathbf{y} - B\alpha \right)^T W^{-1} \left( \mathbf{y} - B\alpha \right) \right] \]

Hence eq. (12), can be rewritten as:

\[ \frac{\partial}{\partial \alpha} \ln (B\alpha, \theta W) = - \frac{1}{2\theta} \frac{\partial}{\partial \alpha} \left[ tr \left( \left( \mathbf{y} - B\alpha \right)^T W^{-1} \left( \mathbf{y} - B\alpha \right) \right) \right] = 0 \] (13)

Since \( \left( \mathbf{y} - B\alpha \right)^T W^{-1} \left( \mathbf{y} - B\alpha \right) \) is a scalar, then its value equal to \( tr \left[ \left( \mathbf{y} - B\alpha \right)^T W^{-1} \left( \mathbf{y} - B\alpha \right) \right] \), so that Eq. (13) can be rewritten as:

\[
\frac{\partial}{\partial \alpha} \left( \mathbf{y}^T W^{-1} \mathbf{y} - \mathbf{2} \alpha^T B^T W^{-1} \mathbf{y} + \alpha^T B^T W^{-1} B \alpha \right) = 0 \\
2B^T W^{-1} B \alpha = 2B^T W^{-1} \mathbf{y} \\
\hat{\alpha} = \left( B^T W^{-1} B \right)^{-1} B^T W^{-1} \mathbf{y} \] (14)

Furthermore, the estimation of \( \hat{f}(\mathbf{x}) \) can be obtained by using Eq. (14), as follows:

\[ \hat{f}(\mathbf{x}) = B \hat{\alpha} = B \left( B^T W^{-1} B \right)^{-1} B^T W^{-1} \mathbf{y} = H \mathbf{y} \] (15)

where \( H = B \left( B^T W^{-1} B \right)^{-1} B^T W^{-1} \). It implies that \( \hat{f}(\mathbf{x}) \) is linear in observation \( \mathbf{y} \).

The form of \( \hat{f}(\mathbf{x}) \) in Eq. (15) contains weighted \( W \) which is unknown its value, therefore \( W \) must be estimated from the data. In this paper, we use MLE method to obtain \( \hat{W} \) as given in the following theorem 2.

**Theorem 2.**

If the bi-responses nonparametric regression model based on MARS function is expressed in Eq.(8) with its assumption in Eq.(10), then estimation of weighted is:

\[
\hat{W} = \left( \begin{array}{cc}
\left( y_1^T A_{11} y_2^T A_{22} y_2 \right)^\dagger & y_1^T A_{12} y_2^T A_{22} y_2 \frac{1}{T} \\
\left( y_2^T A_{21} y_1^T A_{11} y_1 \right)^\dagger & y_2^T A_{22} y_1^T A_{11} y_1 \frac{1}{T}
\end{array} \right) \otimes I_n \] (16)

where: \( A_{1i} = (I_i - H_i)^T (I_i - H_i); \ A_{ii} = (I_i - H_i)^T (I_i - H_i) \)

\( H_i = B_i \left( B_i^T B_i \right)^{-1} B_i^T \); \( H_s = B_s \left( B_s^T B_s \right)^{-1} B_s^T \), for \( i \neq s = 1, 2 \).

**Proof:**

Based on assumption in Eq. (10), then the likelihood function of \( \mathbf{y} \) in Eq. (8), can be written as follows:
\[ L(f(x), \theta W) = (2\pi)^{-n} |(\Sigma \otimes I_n)|^{-1/2} \det \left[ \frac{1}{2} (\Sigma \otimes I_n)^{-1} \begin{pmatrix} y_1 - f_1(x) \\ y_2 - f_2(x) \end{pmatrix}^T \begin{pmatrix} y_1 - f_1(x) \\ y_2 - f_2(x) \end{pmatrix} \right] \]

and the log of the likelihood function is:
\[
\ell n(f(x), \theta W) = -n \log(2\pi) - \frac{1}{2} \log |\Sigma \otimes I_n| - \frac{1}{2} \operatorname{tr} \left[ (\Sigma \otimes I_n)^{-1} \begin{pmatrix} y_1 - f_1(x) \\ y_2 - f_2(x) \end{pmatrix}^T \begin{pmatrix} y_1 - f_1(x) \\ y_2 - f_2(x) \end{pmatrix} \right] (17)\]

Using theorem in Gupta et al. [14], we obtain:
\[
\left| (\Sigma \otimes I_n) \right| = |\Sigma|^n |I_n|^2 = |\Sigma|^n \tag{18}\]

and:
\[
\operatorname{tr} \left[ (\Sigma \otimes I_n)^{-1} \begin{pmatrix} y_1 - f_1(x) \\ y_2 - f_2(x) \end{pmatrix}^T \begin{pmatrix} y_1 - f_1(x) \\ y_2 - f_2(x) \end{pmatrix} \right] = \operatorname{tr} \left[ \Sigma^{-1} \otimes I_n \begin{pmatrix} y_1 - f_1(x) \\ y_2 - f_2(x) \end{pmatrix}^T \begin{pmatrix} y_1 - f_1(x) \\ y_2 - f_2(x) \end{pmatrix} \right] (19)\]

By applying Eq. (18) and (19) to Eq. (17), we get:
\[
\ell n(f(x), \theta W) = -n \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} \operatorname{tr} \left[ \Sigma^{-1} \left( \begin{pmatrix} y_1 - f_1(x) \\ y_2 - f_2(x) \end{pmatrix}^T \begin{pmatrix} y_1 - f_1(x) \\ y_2 - f_2(x) \end{pmatrix} \right) \right] (20)\]

According to Christensen [15], MLE of \( \theta W \) in Eq. (20) can be obtained by using the least square estimation of \( f(x) \) for each response as in Eq. (7), as follows:
\[
\hat{f}_1(x) = B_i \hat{\alpha}_1 = B_i \left( B_i^T B_i \right)^{-1} B_i y_i = H_i y_i \]
\[
\hat{f}_2(x) = B_i \hat{\alpha}_2 = B_i \left( B_i^T B_i \right)^{-1} B_i y_2 = H_2 y_2 \tag{21}\]

where \( H_i = B_i \left( B_i^T B_i \right)^{-1} B_i \) and \( H_2 = B_i \left( B_i^T B_i \right)^{-1} B_i \). As we have known, the least squares estimation doesn’t depend on the variance-covariance matrix. Hence for any value of \( \theta W \), \( \hat{f}(x) \) value in Eq. (21) maximize the log likelihood function in Eq. (20). Consequently, Eq. (20) can be rewritten as:
\[
\ell n(f(x), \theta W) = -n \log(2\pi) - \frac{n}{2} \log |\Sigma| - \frac{1}{2} \operatorname{tr} \left[ \Sigma^{-1} \left( \begin{pmatrix} y_1 - \hat{f}_1(x) \\ y_2 - \hat{f}_2(x) \end{pmatrix}^T \begin{pmatrix} y_1 - \hat{f}_1(x) \\ y_2 - \hat{f}_2(x) \end{pmatrix} \right) \right] = -n \log(2\pi) - \frac{n}{2} \log |\Sigma| - \frac{1}{2} \operatorname{tr} \left[ \Sigma^{-1} \left( \begin{pmatrix} y_1 - H_i y_i \\ y_2 - H_2 y_2 \end{pmatrix}^T \begin{pmatrix} y_1 - H_i y_i \\ y_2 - H_2 y_2 \end{pmatrix} \right) \right] \]

Setting the partial derivative of \( \ell n(\hat{f}(x), \theta W) \) with respect to \( \sigma_{ij} \), \( (i, j = 1, 2) \) equal to zero, we obtain:
\[
\frac{\partial}{\partial \sigma_g} \ln(\hat{f}(\chi), \theta W) = \frac{n}{2} \left( \frac{\partial}{\partial \sigma_g} \log |\Sigma| \right) \ldots
\]

\[
-\frac{1}{2} \frac{\partial}{\partial \sigma_g} tr \left( \Sigma^{-1} \left[ \begin{array}{cc}
(y_i - H_1 y_i) & (y_i - H_1 y_i) \\
(y_i - H_2 y_i) & (y_i - H_2 y_i)
\end{array} \right] \right) = 0
\]  

(22)

To find the partial derivative, we use proposition in Christensen [16], as follows:

\[
\frac{\partial}{\partial \sigma_g} \Sigma^{-1} = -\Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_g} \Sigma^{-1}
\]  

(23)

\[
\frac{\partial}{\partial \sigma_g} \log |\Sigma| = tr \left[ \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_g} \right]
\]  

(24)

and:

\[
\frac{\partial}{\partial \sigma_g} tr \left[ \Sigma^{-1} \right] = tr \left[ \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_g} \right]
\]  

(25)

Thus, by applying Eq. (23), (24) and (25) to Eq. (22), we get:

\[
\frac{\partial}{\partial \sigma_g} \ln(\hat{f}(\chi), \theta W) = -\frac{n}{2} tr \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_g} \right)
\]

(26)

Furthermore, the estimation of \( \theta W \) and \( W \) can be obtained by using Eq. (3) and (26), as follows:

\[
\theta \hat{W} = \frac{1}{n} \left( \begin{array}{cc}
y_i^T A_1 y_i & y_i^T A_2 y_i \\
y_i^T A_2 y_i & y_i^T A_3 y_i
\end{array} \right) \otimes I_n
\]  

(27)

\[
\hat{W} = \left( \begin{array}{cc}
y_i^T A_1 y_i / (y_i^T A_2 y_i) & y_i^T A_2 y_i / (y_i^T A_3 y_i) \\
y_i^T A_2 y_i / (y_i^T A_3 y_i) & y_i^T A_3 y_i / (y_i^T A_1 y_i)
\end{array} \right) \otimes I_n
\]  

(28)

where:

\[
A_i = (I_i - H_i) y_i, \quad A_s = (I_s - H_s) y_s, \quad H_\ell = B_\ell \left( B_\ell^T B_\ell \right)^{-1} B_\ell^T
\]

and

\[
H_\ell = B_\ell \left( B_\ell^T B_\ell \right)^{-1} B_\ell^T, \quad \text{for} \quad \ell \neq s = 1, 2.
\]
In addition, the estimator properties of bi-responses nonparametric regression model with MARS is given in the following lemma.

**Lemma.**
If estimation of \( \hat{f}(x) \) and \( W \) are respectively given in Eq. (15) and (27), then the properties of \( \hat{f}(x) \) is a bias estimator.

**Proof:**
Based on Eq. (15), then the expectation value of \( \hat{f}(x) \) can be written as follows:

\[
E[\hat{f}(x)] = E(H_y) = H.E(y)
\]

\[
= H.E(f(x) + \varepsilon)
\]

\[
= H.f(x)
\]

(29)

Since \( E[\hat{f}(x)] \neq f(x) \) then \( f(x) \) is a bias estimator.

The bias value in Eq. (29) can be minimized by selecting the optimal basis function \( B_{opt} \) in \( H \) matrix, in order to obtain the best estimator which is closest to the actual value. According to Friedman [6], selection of \( B_{opt} \) in MARS can be obtained by using Generalized Cross Validation (GCV) method.

**5. Conclusion**

Based on the results of this study which is presented on the previous part, we can be drawn several conclusions as follows:
1. The equation form of bi-responses nonparametric regression model using MARS function as \( y = B\alpha + \varepsilon \).
2. Estimation from its model can be obtained by maximum likelihood estimation (MLE) method, thus obtained:

\[
\hat{f}(x) = B(B^TW^{-1}B)^{-1}B^W^{-1} = H_y^*, \quad \text{where} \quad H = B(B^TW^{-1}B)^{-1}B^W^{-1}.
\]
3. The estimation of \( f(x) \) contains weighted i.e. variance-covariance matrix \( W \) which is unknown its value, therefore it must be estimated from the data. By using MLE, we obtain estimation from \( W \) as:

\[
\hat{W} = \left( \begin{array}{cc}
(y_1^TA_{11}y_1^T A_{21}y_1^T y_2^T A_{12}y_2^T y_1^T A_{11}y_1^T y_2^T A_{12}y_2^T y_2^T A_{22}y_2^T)
\end{array} \right) \otimes I_n, \quad \text{where}:
\]

\[A_s = (I - H_s)^T (I - H_s), \quad A_s = (I - H_s)^T (I - H_s), \quad H_s = B_s (B_s^T B_s)^{-1} B_s^T \]

and \( H_s = B_s (B_s^T B_s)^{-1} B_s^T \), for \( s = 1, 2 \).
4. The estimator properties from its model is a bias estimator and linear in observation \( y \).
Acknowledgements. The first author would like to thanks BPS-Statistics Indonesia for the Ph.D. grant through Human Resources Development Project in cooperation with Sepuluh Nopember Institute of Technology Surabaya.

References


Received: January 21, 2015; Published: March 27, 2015