5th ICRIEMS Proceedings
Published by Faculty Of Mathematics And Natural Sciences
Yogyakarta State University, ISBN 978-602-74529-3-0

The Estimation of Binary Nonparametric Regression Model based on The Kernel Estimator by Generalized Additive Models Method

Suliyanto[1,a)] , Marisa Rifada[1,b)] , Eko Tjahjono[1,c)] and Sediono[1,d)]

[1]*Department of Mathematics, Airlangga University, Indonesia*

a)yanfit@yahoo.com
b) marisa.rifada@fst.unair.ac.id
c)eko-t@fst.unair.ac.id
d)sediono101@gmail.com

Abstract. Binary regression model is a regression model with response variable is categorical that consists of two categories, while the predictor variables are continuous or categorical. In a binary regression model, the effect of predictor variables on the response variable is declared as logit function. If the logit function is assumed to be linear in the parameters of the models, it is included in the Generalized Linear Models (GLM) and estimated by parametric approach. Whereas if the logit function is assumed not to be a certain form or there is not an assumption based on theory or experience of the past, it is called binary nonparametric regression model and estimated by nonparametric approach. In this study, we used Generalized Additive Models (GAM) method associated with nonparametric regression which assumes that the nonparametric regression function is expressed as the sum of nonparametric regression functions were assumed that function form is unknown with the known link function. The aim of this study is estimating each nonparametric regression function of each predictor variable using Kernel estimator. The result of estimating binary nonparametric regression model by GAM method based on kernel estimator is obtained by used Local Scoring algorithms. This algorithm consists of two loops that are Scoring step (outer loop) is iterated until the average value of deviance convergent and weighted Backfitting step (inner loop) is iterated until the average value Residual Sum of Squares (RSS) convergent.

INTRODUCTION

Statistical methods used to analyze the relationship between the response variable categorical scale with predictor variables categorical or continuous scale is logistic regression. Binary logistic regression was logistic regression with response variable consisting of two categories [1]. There are two ways to approach the regression model, i.e. parametric and nonparametric approach. Parametric approach assumes that the regression model for each individual observation has the same parameters, while the nonparametric approach assumes that not all individuals have the same parameters.

According [2] one way to estimate the nonparametric regression model is to use local likelihood estimation method. Particularly in binary nonparametric logistic regression model or known by the nonparametric regression model with ~~a~~ binary response variable. To estimate the model used local likelihood logit estimation method [2] or Generalized Additive Models (GAM) [3]. The local likelihood logit estimation method includes logit model~~s~~ with kernel weighting function for an infinite bandwidth value [2]. The GAM method assumes that the logit function is expressed as the sum of nonparametric regression functions that are unknown from each predictor variable with the known logit link function. The GAM method is an extension of the method of Generalized Linear Models (GLM) by replacing the linear function within the parameters of the GLM with the sum of the unknown nonparametric regression functions of each predictor variable. The distribution of the response variable on the GAM is not limited to the Normal distribution, but included in the exponential family [3].

In this study, the binary nonparametric regression model assumes that the response variable is binary categorical, whereas the predictor variables are continuous with the logit function is expressed as the sum of nonparametric regression functions which unknown function form. Each nonparametric regression function of each predictor variable is estimated using Kernel estimator. The algorithm used to obtain the estimates of nonparametric regression functions is the Local Scoring algorithm that very suitable usage in additive nonparametric regression model with the distribution of the response variable was included in the exponential family members, one of which is a Bernoulli distribution [3]. This algorithm consists of two loops that are Scoring step (outer loop) is iterated until the average value of deviance convergent and weighted Backfitting step (inner loop) is iterated until the average value of the Residual Sum of Squares (RSS) convergent.

METHODS

Generalized Additive Models

The GAM method is one method of nonparametric regression approach introduced by Hastie and Tibshirani [3]. The GAM method associated with nonparametric regression assumes that the nonparametric regression function is expressed as the sum of nonparametric regression functions of each component x which unknown form with the link function G is known, i.e :

$$G\left(E\left(Y \mid X = x\right)\right) = \sum_{j=1}^{p} f_j(x_j) \tag{1}$$

with $f_j(x_j)$ is a nonparametric regression function of the *j-th* predictor variable.

Binary Nonparametric Regression Model

Binary nonparametric regression model assumes that the response variable is binary categorical (0,1), whereas the predictor variables are continuous with the logit function is expressed as the sum of nonparametric regression functions $f_j(x_{ji})$ of the *j-th* predictor variable to *i-th* observation with unknown function form and known link function G. If the selected logit link function G, then using (1) the binary nonparametric regression model can be obtained below :

$$\ln\left[\frac{E\left(Y \mid X = x_i\right)}{1 - E\left(Y \mid X = x_i\right)}\right] = \sum_{j=1}^{p} f_j(x_{ji}) \tag{2}$$

where $E\left(Y \mid X = x_i\right) = P\left(Y = 1 \mid X = x_i\right) = p_i$ is the probability of success in *i-th* observation corresponding to the predictor variable X.

Kernel Functions

According to Hardle [4], the Kernel function with bandwidth h is defined as :

$$K_h\left(x\right) = \frac{1}{h} K\left(\frac{x}{h}\right), \text{ for } -\infty < x < \infty,\ h > 0 \tag{3}$$

The Kernel function K has the properties as follows : $K(x) \geq 0$ for all x, $\int_{-\infty}^{\infty} K(x)\,dx = 1$, $\int_{-\infty}^{\infty} xK(x)\,dx = 0$, and $\int_{-\infty}^{\infty} x^2 K(x)\,dx = t^2 > 0$. There are several types of the Kernel functions, i.e :

1. Uniform : $K(x) = \frac{1}{2} I(|x| \leq 1)$
2. Triangle : $K(x) = (1 - |x|) I(|x| \leq 1)$
3. Gaussian : $K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \; -\infty < x < \infty$

with I is the indicator function.

Optimal Bandwidth Parameters Selection

Selection of bandwidth h is very important in getting the estimator of regression function based on the nonparametric approach. The Bandwidth is the controller of the balance between the smooth of function to the data, if h is very small then the estimation of function obtained will be very rough and heading to the data, whereas if h is very large then the estimation of function obtained will be very smooth and heading to the average of the response variable. Therefore, selecting the optimal value h is expected. There are two methods to obtain optimal h, i.e the Cross Validation (CV) method and the Generalized Cross Validation (GCV) method [5]. The method chosen in this study is GCV method, because GCV method is more asymptotically and does not contain population parameter. If a criterion for h limited to the linear estimator class, then for each h there is a matrix $A(h)$ sized $n \times n$ to obtain the estimation of nonparametric regression function as follows :

$$\hat{f}(x) = A(h)Y \tag{4}$$

The optimal value h in (3) is obtained by minimizing the GCV as follows:

$$GCV(h) = \frac{MSE(h)}{\left(n^{-1} tr\left[I - A(h)\right]\right)^2} \tag{5}$$

where n is size of observations, I is a identity matrix sized $n \times n$, and $MSE(h) = n^{-1} \sum_{i=1}^{n} \left(y_i - \hat{f}(x_i)\right)^2$.

RESULTS AND DISCUSSION

Estimation of Binary Nonparametric Regression Model

Based on equation (2) can be obtained the probability of success if x_i known, i.e :

$$\pi_i = \frac{\exp\left[\sum_{j=1}^{p} f_j(x_{ji})\right]}{1 + \exp\left[\sum_{j=1}^{p} f_j(x_{ji})\right]} \tag{6}$$

Function $f_j(x_{ji})$ which unknown form is estimated by Kernel estimator approach. According to Demir and Toktamis [6], Kernel estimator for the single predictor variable is as follows:

$$\hat{f}(x) = \frac{\sum_{i=1}^{n} K_h(x - x_i) y_i}{\sum_{i=1}^{n} K_h(x - x_i)} \tag{7}$$

with h is the optimal bandwidth and K is a Kernel function. Equation (7) is known as Kernel estimator Nadaraya - Watson for the single predictor variable. Since the predictor variable *j-th* has the bandwidth h_j, then from equation (7) is obtained Kernel estimator Nadaraya – Watson for n observations with predictor variable *j-th* to the observations *i-th* is :

$$\hat{f}_j(x_{ji}) = \frac{\sum_{k=1}^{n} K_{h_j}(x_{ji} - x_{jk}) y_k}{\sum_{k=1}^{n} K_{h_j}(x_{ji} - x_{jk})}, \; i = 1, 2, \dots, n \tag{8}$$

Equation (8) can be expressed as

$$\hat{f}_j(x_{ji}) = \frac{K_{h_j}(x_{ji} - x_{j1}) y_1 + K_{h_j}(x_{ji} - x_{j2}) y_2 + \dots + K_{h_j}(x_{ji} - x_{jn}) y_n}{K_{h_j}(x_{ji} - x_{j1}) + K_{h_j}(x_{ji} - x_{j2}) + \dots + K_{h_j}(x_{ji} - x_{jn})} = \frac{L' W_{ji}(h_j) Y}{L' W_{ji}(h_j) L} \tag{9}$$

where

$$W_{ji}(h_j) = diag\left[K_{h_j}(x_{ji} - x_{j1}), K_{h_j}(x_{ji} - x_{j2}), \dots, K_{h_j}(x_{ji} - x_{jn})\right] \tag{9.a}$$

$Y = (y_1, y_2, \dots, y_n)'$ and $L = (1, 1, \dots, 1)'$

Based on equation (9), we get

$$\begin{bmatrix} \hat{f}_j(x_{j1}) \\ \hat{f}_j(x_{j2}) \\ \vdots \\ \hat{f}_j(x_{jn}) \end{bmatrix} = \begin{bmatrix} \frac{L' W_{j1}(h_j) Y}{L' W_{j1}(h_j) L} \\ \frac{L' W_{j2}(h_j) Y}{L' W_{j2}(h_j) L} \\ \vdots \\ \frac{L' W_{jn}(h_j) Y}{L' W_{jn}(h_j) L} \end{bmatrix} = \begin{bmatrix} \frac{L' W_{j1}(h_j)}{L' W_{j1}(h_j) L} \\ \frac{L' W_{j2}(h_j)}{L' W_{j2}(h_j) L} \\ \vdots \\ \frac{L' W_{jn}(h_j)}{L' W_{jn}(h_j) L} \end{bmatrix} Y \; ; \; j = 1, 2, \dots, p \tag{10}$$

Equation (10) can be expressed as follows:

$$\hat{f}_j(x_j) = A(h_j) Y \tag{11}$$

where

$$\hat{f}_j(x_j) = \begin{bmatrix} \hat{f}_j(x_{j1}) \\ \hat{f}_j(x_{j2}) \\ \vdots \\ \hat{f}_j(x_{jn}) \end{bmatrix} \quad (11.a)$$

$$A(h_j) = \begin{bmatrix} \frac{L' W_{j1}(h_j)}{L' W_{j1}(h_j) L} \\ \frac{L' W_{j2}(h_j)}{L' W_{j2}(h_j) L} \\ \vdots \\ \frac{L' W_{jn}(h_j)}{L' W_{jn}(h_j) L} \end{bmatrix} \quad (11.b)$$

In this study, the binary nonparametric regression model has the response variable which binary categorical (0,1) and distributed Bernoulli. We estimate the binary nonparametric regression model based on Kernel estimator used the Local scoring algorithm because this algorithm very suitable usage in additive nonparametric regression model with the distribution of the response variable was included in the exponential family members, one of which is a Bernoulli distribution. This algorithm consists of two loops that are Scoring step (outer loop) is iterated until the average value of deviance convergent and weighted Backfitting step (inner loop) is iterated until the average value of the Residual Sum of Squares (RSS) convergent. Scoring step is done iteratively by determining the adjusted value of the response variable (z) which is formulated as follows :

$$z_i = m_i^{(s)} + \left(y_i - \pi_i^{(s)}\right)\left(\frac{\partial m_i}{\partial \pi_i}\right)_{(s)}, \ i = 1, 2, ..., n \quad (12)$$

where $m_i^{(s)} = \sum_{j=1}^{p} f_j^{(s)}(x_{ji}) = \ln\left(\frac{p_i^{(s)}}{1 - p_i^{(s)}}\right)$, s = 0,1,2, ... (12.a)

The weighting matrix B is a diagonal matrix with the main diagonal elements are :

$$b_i = \left(\frac{\partial \pi_i}{\partial m_i}\right)_{(s)}^2 \left(V_i^{(s)}\right)^{-1}, \ i = 1, 2, ..., n \quad (13)$$

Since the response variable Y has Bernoulli distribution, then we have :

$$V_i = \text{Var}(Y_i) = \pi_i\left(1 - \pi_i\right) \quad (14)$$

Based on equation (12.a) we get :

$$\frac{\partial m_i}{\partial \pi_i} = \frac{1}{\pi_i} + \frac{1}{1-\pi_i} = \frac{1}{\pi_i(1-\pi_i)} \quad (15)$$

So, from equation (15) we can obtain :

$$\frac{\partial \pi_i}{\partial m_i} = \pi_i(1-\pi_i) \quad (16)$$

Then, by substituting equation (15) into equation (12) we have :

$$z_i = m_i^{(s)} + \frac{\left(y_i - \pi_i^{(s)}\right)}{\pi_i^{(s)}\left(1-\pi_i^{(s)}\right)} \tag{17}$$

Next, by substituting equations (14) and (16) into equation (13) we obtain :

$$b_i = \pi_i(1-\pi_i) \tag{18}$$

Subsequently to step weighted Backfitting is done iteratively by determining the estimation of nonparametric regression functions $\hat{f}_j(x_j)$ in the model which is formulated as follows:

$$\hat{f}_j(x_j) = A(h_j)\left\{z - \sum_{k=1}^{j-1} \hat{f}_k^{(s+1)}(x_k) - \sum_{k=j+1}^{p} \hat{f}_k^{(s)}(x_k)\right\} \quad \text{, j = 1, 2, ..., p} \tag{19}$$

with $A(h_j)$ according to the equation (11.b) and s = 0,1,2, ...

Finally, we obtain estimation of binary nonparametric regression model based on the Kernel estimator using Local scoring algorithm as follows:

$$\ln\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = \sum_{j=1}^{p} \hat{f}_j(x_{ji})$$

$$= \sum_{j=1}^{p} \frac{L'W_{ji}(h_j)}{L'W_{ji}(h_j)L}\left\{z - \sum_{k=1}^{j-1} \hat{f}_k^{(s+1)}(x_k) - \sum_{k=j+1}^{p} \hat{f}_k^{(s)}(x_k)\right\} \tag{20}$$

Algorithm for Estimating Binary Nonparametric Regression Model

To estimate the binary nonparametric regression model based on the Kernel estimator by the GAM method we used the following algorithm :

(1) The algorithm for determining the optimal bandwidth value for each predictor variable.

1. Input pair data $(y_i, x_{1i}, x_{2i}, ..., x_{pi})$; $i = 1, 2, ..., n$
2. Defining the Kernel function used is Gaussian Kernel function
3. Determining the initial bandwidth value (h_j)
4. Determining a diagonal weighting matrix $W_{ji}(h_j)$ such as the equation (9.a)
5. Determining matrix $A(h_j)$ such as the equation (11.b)
6. Calculating the value of $\hat{f}_j\left(x_j\right) = A(h_j)Y$
7. Calculating the value of $MSE(h_j) = n^{-1}\sum_{i=1}^{n}(y_i - \hat{f}_j(x_{ji}))^2$
8. Calculating the value of $GCV(h_j) = \dfrac{MSE(h_j)}{(n^{-1}tr[I - A(h_j)])^2}$
9. Repeat steps 4 to 8 until we get minimum value of $GCV(h_j)$ for optimal bandwidth (h_j).

(2) The algorithm for initial estimation of nonparametric regression function $\hat{f}_j\left(x_j\right)$ for each predictor variable.

1. Input pair data $(y_i, x_{1i}, x_{2i}, ..., x_{pi})$; $i = 1, 2, ..., n$
2. Defining the kernel function used is Gaussian kernel function

3. Input the optimal bandwidth value (h_j) obtained from the algorithm (1).
4. Determining a diagonal weighting matrix $W_{ji}(h_j)$ such as the equation (9.a)
5. Determining matrix $A(h_j)$ such as the equation (11.b)
6. Calculating the value of $\hat{f}_j\left(x_j\right) = A(h_j)Y$

(3) The Local scoring algorithm to estimate the binary nonparametric regression model.

1. Input pair data $(y_i, x_{1i}, x_{2i}, ..., x_{pi})$; $i = 1, 2, ..., n$
2. Input the initial estimation value of $\hat{f}_j^{(0)}(x_j)$; $j = 1, 2, ..., p$ obtained from the algorithm (2)
3. Iterating Scoring step (outer loop) as follows:
   a. Determining the adjusted value of the response variable (z) with elements of row i-th is

$$z_i = m_i^{(s)} + \frac{\left(y_i - \pi_i^{(s)}\right)}{\pi_i^{(s)}\left(1-\pi_i^{(s)}\right)} \quad , \quad i = 1,2,...,n$$

   with $m_i^{(s)} = \sum_{j=1}^{p} \hat{f}_j^{(s)}(x_{ji})$, $\pi_i^{(s)} = \frac{\exp(m_i^{(s)})}{1+\exp(m_i^{(s)})}$ and determining the weighting matrix (B) in the form of a diagonal matrix with the main diagonal elements are

$$b_i = \pi_i^{(s)}\left(1-\pi_i^{(s)}\right) \quad , \quad i = 1, 2, \ldots, n$$

   b. Iterating weighted Backfitting step (inner loop) as follows:

   (i) For the initial iteration ($s = 0$), Defining $\hat{m}_j^{(s)}(X_j) = \hat{m}_j^{(r)}(X_j)$ and $z = \left(z_1, z_2, \ldots, z_n\right)^T$

   (ii) Estimating the nonparametric regression functions in the model for $j = 1, 2, \ldots, p$, that is

$$\hat{f}_j^{(s+1)}(X_j) = A(h_j)\left\{z - \sum_{k=1}^{j-1} \hat{f}_k^{(s+1)}(X_k) - \sum_{k=j+1}^{p} \hat{f}_k^{(s)}(X_k)\right\}$$

   (iii) Calculating the average value of the weighted Residual Sum of Squares (RSS)

$$Avg(RSS)^{(s+1)} = \frac{1}{n}\left\{\left(z^{(s+1)} - m^{(s+1)}\right)^T B^{(s+1)}\left(z^{(s+1)} - m^{(s+1)}\right)\right\}$$

   (iv) Repeat steps (ii) to (iii) for $s = s+1$ until the average value of RSS convergent, i.e : $abs\left(Avg(RSS)^{(s+1)} - Avg(RSS)^{(s)}\right) < \varepsilon$, for $\varepsilon = 0.0001$

   c. Defining $\hat{f}_j^{(r+1)}(x_j) = \hat{f}_j^{(s+1)}(x_j)$ and then determining the average value of deviance :

$$Avg\left(D(y_i;\pi_i)\right)^{(r+1)} \approx \frac{-2}{n}\sum_{i=1}^{n}\left\{y_i \ln \pi_i^{(r+1)} + (1-y_i)\ln(1-\pi_i^{(r+1)})\right\}$$

   d. Repeat step a to step c for $r = r+1$ until the average value of deviance convergent, i.e : $abs\left(Avg\left(D(y_i;\pi_i)\right)^{(r+1)} - Avg\left(D(y_i;\pi_i)\right)^{(r)}\right) < \varepsilon$, with $\varepsilon = 0.0001$

CONCLUSION

The result of estimating binary nonparametric regression model by GAM method based on Kernel estimator using Local scoring algorithm is obtained as follows :

$$\ln\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right)=\sum_{j=1}^{p}\frac{L'W_{ji}(h_j)}{L'W_{ji}(h_j)L}\left\{z-\sum_{k=1}^{j-1}\hat{f}_k^{(s+1)}(x_k)-\sum_{k=j+1}^{p}\hat{f}_k^{(s)}(x_k)\right\}$$

The local scoring algorithm consists of two loops that are Scoring step (outer loop) is iterated until the average value of deviance convergent and weighted Backfitting step (inner loop) is iterated until the average value of the Residual Sum of Squares (RSS) convergent.

ACKNOWLEDGMENTS

The authors thanks to the Ministry of Research, Technology and Higher Education of the Republic of Indonesia for financial support of this research through The Fundamental Research University Grant 2018. The authors also thank to the anonymous referees for their valuable suggestions which let to the improvement on the manuscript.

REFERENCES

1. A. Agresti, “Categorical Data Analysis”, Second Edition, John Wiley and Sons, New York, 2002.
2. M. Frölich, “ Non-parametric regression for binary dependent variables”, in *Econometrics Journal*, volume 9, pp. 511-540 (2006).
3. T. J. Hastie, and R. J. Tibshirani, “ Generalized Additive Models”, Chapman & Hall, London, 1990.
4. W. Hardle, “ Applied Nonparametrik Regression”, Cambridge University Press, New York, 1990.
5. R. L. Eubank, “ Spline Smoothing and Nonparametric Regression”, Marcel Deker, New York and Basel, 1988.
6. S. Demir, and O. Toktamis, “On The Adaptive Nadaraya-Watson Kernel Regression Estimators”, in *Hacettepe Journal of Mathematics and Statistics*, volume 39(3), pp 429-437 (2010).