Journal of the Korean Data & Information Science Society
2016, **27**(2), 531–538

http://dx.doi.org/10.7465/jkdi.2016.27.2.531
한국데이터정보과학회지

Geographically weighted kernel logistic regression for small area proportion estimation†

Jooyong Shim[1] · Changha Hwang[2]

[1]Department of Statistics, Inje University
[2]Department of Applied Statistics, Dankook University
Received 15 January 2016, revised 18 February 2016, accepted 22 March 2016

Abstract

In this paper we deal with the small area estimation for the case that the response variables take binary values. The mixed effects models have been extensively studied for the small area estimation, which treats the spatial effects as random effects. However, when the spatial information of each area is given specifically as coordinates it is popular to use the geographically weighted logistic regression to incorporate the spatial information by assuming that the regression parameters vary spatially across areas. In this paper, relaxing the linearity assumption and propose a geographically weighted kernel logistic regression for estimating small area proportions by using basic principle of kernel machine. Numerical studies have been carried out to compare the performance of proposed method with other methods in estimating small area proportion.

Keywords: Geographically weighted regression, kernel machine, logistic regression, mixed effects model, small area estimation.

1. Introduction

Small area is a small subpopulation constituted by specific demographic and socioeconomic group of people, within a larger geographical areas. Small area estimation (SAE) is any of several statistical techniques involving the estimation of parameters for small areas for which there is insufficient sample to yield direct estimates with adequate precision and reliability. The underlying idea of SAE is to use auxiliary data at the small area level in combination with survey data to model the small area parameters of interest. A variety of models has been developed for this purpose. SAE has received considerable attention in recent years due to a growing demand for reliable small area statistics for policy analysis and planning purposes.

The linear mixed effects models have been commonly used to solve SAE (Harville, 1976; Jiang and Lahiri, 2000). These models incorporate area-specific random effect to deal with

---

† This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2014R1A1A2054917).
[1] Adjunct professor, Institute of Statistical Information, Department of Statistics, Inje University, Kimhae 621-749, Korea.
[2] Corresponding author: Professor, Department of Applied Statistics, Dankook University, Gyeonggi-do 448-701, Korea. E-mail: chwang@dankook.ac.kr

between-area variation not explained by the available covariates (Rao, 2003). However these models suffer the drawbacks of not reflecting complicated patterns between the response variable and the covariates (Pfeffermann, 2013).

Geographically weighted regression (GWR) (Brunsdon *et al.*, 1996; Yu and Wu, 2004) is an alternative approach to spatial autoregressive regression (Anselin, 1992), which extends the classical regression model by allowing local parameters to be estimated. GWR incorporates the spatial information by assuming that the regression parameters vary spatially across the geography of interest. In GWR the response variable could not necessarily follow a normal distribution but Bernoulli distribution which leads the geographically weighted logistic regression.

The informative method and the discriminative method are the main methods of classification. Classical linear discriminant analysis (LDA) is the most popular informative method, whereas logistic regression is the most popular discriminative method. The logistic regression is more robust than LDA generally, since less assumptions about the classes are made. An important advantage of logistic regression is that it outputs an estimate of the probability that an object belongs to each of the possible classes. A different approach to discriminative method is the support vector machine (SVM) by Vapnik (1995, 1998) and Least squares SVM (LS-SVM) by Suykens and Vandewalle (1999). For applications of SVM and LS-SVM see Hwang (2014, 2015), Seok (2015) and Shim and Seok (2014). Compared to logistic regression, the main drawback of the SVM and LS-SVM is the absence of probabilistic outputs. The most popular methods for the classification in recent machine learning research are variants on SVM, LS-SVM and boosting, sometimes combined with error-correcting code approaches. Rifkin and Klautau (2004) provide a review. In this paper we focus on a nonlinear kernelized variant of logistic regression incorporating the spatial information.

The rest of this paper is organized as follows. Section 2 describes the mixed effects (kernel) logistic regression for spatial data. Section 3 proposes the geographically weighted kernel logistic regression (GWKLR) for small area proportion estimation. Section 4 and 5 present numerical studies and conclusion, respectively.

2. Mixed effects logistic regression

2.1. Mixed effects logistic regression

Let a finite population of size N be divided into m areas such that $N = \sum_{i=1}^{m} N_i$. Suppose that y_{ij} is the indicator function such that 1 if the jth unit in the ith area belongs to a certain class for $i = 1, 2, \cdots, m$, $j = 1, 2, \cdots, N_i$. We focus on estimating the proportion of a certain class in the ith area, which is given by $p_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}$.

We briefly review a mixed effects logistic regression (MELR) model,

$$\log \frac{p_{ij}}{1 - p_{ij}} = \boldsymbol{\beta}' \boldsymbol{x}_{ij} + b_i + b_0 \text{ for } i = 1, \cdots, m,\ j = 1, \cdots, N_i, \quad (2.1)$$

where $\boldsymbol{x}_{ij}$ is the input vector, b_i can be interpreted as the area effect and p_{ij} is the probability that the jth unit in the ith area belongs to a certain class,

$$p_{ij} = P(y_{ij} = 1 | \boldsymbol{x}_{ij}) = \frac{\exp(\boldsymbol{\beta}' \boldsymbol{x}_{ij} + b_i + b_0)}{1 + \exp(\boldsymbol{\beta}' \boldsymbol{x}_{ij} + b_i + b_0)}. \quad (2.2)$$

Then the estimator of p_i is obtained as follows:

$$\widehat{p}_i = \frac{1}{N_i}\left(\sum_{j\in S_i} y_{ij} + \sum_{j\in R_i} \widehat{y}_{ij}\right), \tag{2.3}$$

where $\widehat{y}_{ij}$ is 0 or 1, S_i and R_i represent the sampled and non-sampled sets of units in the ith area, respectively, with $U_i = S_i \cup R_i$. Here, U_i is the set of whole units in the ith area, and $\widehat{y}_{ij}$ presents the predicted value of response for population unit $j \in R_i$. We actually utilize the biased-adjusted estimator (Salvati *et al.*, 2011) given by

$$\widehat{p}_i = \frac{1}{N_i}\left(\sum_{j\in S_i} y_{ij} + \sum_{j\in R_i} \widehat{y}_{ij} + \frac{N_i - n_i}{n_i}\sum_{j\in S_i}(y_{ij} - \widehat{y}_{ij})\right), \tag{2.4}$$

where n_i is the number of units in S_i and $\widehat{y}_{ij}$ is the predicted value of response for population unit $j \in R_i$ or $j \in S_i$. Thus $\widehat{y}_{ij}$ in (2.3) and (2.4) are obtained by using the round-off function [ ] such as $\widehat{y}_{ij} = [\widehat{p}_{ij}]$, where

$$\widehat{p}_{ij} = \frac{\exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{x}_{ij} + \widehat{b}_i + \widehat{b}_0)}{1 + \exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{x}_{ij} + \widehat{b}_i + \widehat{b}_0)} \tag{2.5}$$

where $\boldsymbol{x}_{ij}$ is the input vector for population unit $j \in R_i$ or $j \in S_i$.

With sampled units the estimate of $(\boldsymbol{\beta}, b_i, b_0)$ is obtained by minimizing the negative log-likelihood given by

$$L(\boldsymbol{\beta}, b_i, b_0) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}[-y_{ij}(\boldsymbol{\beta}'\boldsymbol{x}_{ij} + b_i + b_0) + \log(1 + \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij} + b_i + b_{0)})]. \tag{2.6}$$

2.2. Mixed effects kernel logistic regression

We now briefly review a mixed effects kernel logistic regression (MEKLR) model developed by Shim and Hwang (2014). Using the nonlinear feature mapping function ϕ, we make the penalized negative log-likelihood with sampled units given by

$$L(\boldsymbol{\omega}, b_i, b_0) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}[-y_{ij}(\boldsymbol{\omega}'\phi(\boldsymbol{x}_{ij}) + b_i + b_0) + \log(1 + \exp(\boldsymbol{\omega}'\phi(\boldsymbol{x}_{ij}) + b_i + b_0))] + \frac{\lambda}{2}\boldsymbol{\omega}'\boldsymbol{\omega}, \tag{2.7}$$

where $\lambda > 0$ is a penalty parameter. Here $\phi(\cdot) : R^d \to R^{d_f}$ maps the input space to the higher dimensional feature space where the dimension d_f is defined in an implicit way. We know that $\phi(\boldsymbol{x}_1)'\phi(\boldsymbol{x}_2) = K(\boldsymbol{x}_1, \boldsymbol{x}_2)$, where $K(\cdot,\cdot)$ is a kernel function obtained from the application of Mercer's conditions (1909). According to Kimeldorf and Wahba (1971), $\boldsymbol{\omega}'\phi(\boldsymbol{x}_{ij})$ can be written over sampled units as follows:

$$\boldsymbol{\omega}'\phi(\boldsymbol{x}_{ij}) = K(\boldsymbol{x}_{ij}, \boldsymbol{x})\boldsymbol{\alpha}$$

for some weight vector $\boldsymbol{\alpha}$ and $\boldsymbol{x} = \{\boldsymbol{x}_{ij}\}_{i=1,j=1}^{m,n_i}$.

The estimate of $(\boldsymbol{\alpha}, b_i, b_0)$ is obtained by minimizing the penalized negative log-likelihood given by

$$L(\boldsymbol{\alpha}, b_i, b_0) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\{-y_{ij}(K_{ij}\boldsymbol{\alpha}+b_i+b_0)+\log(1+\exp(K_{ij}\boldsymbol{\alpha}+b_i+b_0))\}+\frac{\lambda}{2}\boldsymbol{\alpha}'K\boldsymbol{\alpha}, \quad (2.8)$$

where $K_{ij} = K(\boldsymbol{x}_{ij}, \boldsymbol{x})$ and $K = K(\boldsymbol{x}, \boldsymbol{x})$. Thus $\widehat{p}_{ij}$ is obtained as

$$\widehat{p}_{ij} = \frac{\exp(K(\boldsymbol{x}_{ij}, \boldsymbol{x})\widehat{\boldsymbol{\alpha}} + \widehat{b}_i + \widehat{b}_0)}{1 + \exp(K(\boldsymbol{x}_{ij}, \boldsymbol{x})\widehat{\boldsymbol{\alpha}} + \widehat{b}_i + \widehat{b}_0)} \text{ for } j \in R_i \text{ or } j \in S_i. \quad (2.9)$$

3. Geographically weighted kernel logistic regression

Given sampled data set $\{\boldsymbol{x}_{ij}, \boldsymbol{u}_i, y_{ij}\}_{i=1,j=1}^{m,n_i}$ with each input vector $\boldsymbol{x}_{ij} \in R^d$, $\boldsymbol{u}_i \in R^2$ (longitude, latitude) and corresponding response $y_{ij} \in \{0, 1\}$, we assume that y_{ij} given $\boldsymbol{x}_{ij}$ at location $\boldsymbol{u}_k$ follows Bernoulli distribution (p_{ij}^k), where the probability p_{ij}^k depends on ($\boldsymbol{x}_i$, $\boldsymbol{u}_k$) given by

$$p_{ij}^k = \frac{\exp(\boldsymbol{\omega}_k'\phi(\boldsymbol{x}_{ij}) + b_k)}{1 + \exp(\boldsymbol{\omega}_k'\phi(\boldsymbol{x}_{ij}) + b_k)}. \quad (3.1)$$

Here $(\boldsymbol{\omega}_k, b_k)$ is estimated from the penalized negative log- likelihood given by

$$L(\boldsymbol{\omega}_k, b_k) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} v_{ki}[-y_{ij}(\boldsymbol{\omega}_k'\phi(\boldsymbol{x}_{ij})+b_k)+\log(1+\exp(\boldsymbol{\omega}_k'\phi(\boldsymbol{x}_{ij})+b_k)]+\frac{\lambda}{2}\boldsymbol{\omega}_k'\boldsymbol{\omega}_k, \quad (3.2)$$

where $\lambda > 0$ is a penalty parameter, the weight v_{ki} at kth location $\boldsymbol{u}_k$ is taken as a function of the distance from $\boldsymbol{u}_k$ to other locations such that $v_{ki} = \exp(-d_{ki}/h)$, d_{ki} is the distance between location $\boldsymbol{u}_k$ and $\boldsymbol{u}_i$ and $h > 0$ is the bandwidth parameter. According to Kimeldorf and Wahba (1971), the penalized negative log-likelihood (3.2) can be rewritten as follows:

$$L(\boldsymbol{\alpha}_{.k}, b_k) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} v_{ki}[-y_{ij}(K_{ij}\boldsymbol{\alpha}_{.k}+b_k)+\log(1+\exp(K_{ij}\boldsymbol{\alpha}_{.k}+b_k)]+\frac{\lambda}{2}\boldsymbol{\alpha}_{.k}'K\boldsymbol{\alpha}_{.k}. \quad (3.3)$$

Minimizing (3.3) yields the estimate of parameter vector $(\boldsymbol{\alpha}_{.k}, b_k)$, but not in an explicit form, which leads to the iterative reweighted least squares (IRWLS) procedure. At each iteration, the parameter vector $(\boldsymbol{\alpha}_{.k}, b_k)$ is estimated as follows:

$$\begin{pmatrix} \boldsymbol{\alpha}_{.k} \\ b_k \end{pmatrix} = (K_1'W_kV_kK_1 + \lambda K_0)^{-1}K_1'W_kV_k\boldsymbol{y}_{k.}^*, \quad (3.4)$$

where V_k is a diagonal matrix of v_{ki}, W_k is a diagonal matrix composed of $\widehat{p}_{ij}^k(1 - \widehat{p}_{ij}^k)$, $K_1 = (K, \mathbf{1})$, $K_0 = \begin{pmatrix} K & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix}$ and $\boldsymbol{y}_k^* = W_k^{-1}(\boldsymbol{y} - \widehat{\boldsymbol{p}}^k) + K_1\begin{pmatrix} \boldsymbol{\alpha}_{.k} \\ b_k \end{pmatrix}$ is the working response vector. The convergence of IRWLS procedure (3.4) is guaranteed by Perez-Cruz, *et al.* (2005). Finally $\widehat{p}_{ij} = \widehat{p}_{ij}^i$ is obtained as follows:

$$\widehat{p}_{ij} = \frac{\exp(K(\boldsymbol{x}_{ij}, \boldsymbol{x})\widehat{\boldsymbol{\alpha}}_{.i} + \widehat{b}_i)}{1 + \exp(K(\boldsymbol{x}_{ij}, \boldsymbol{x})\widehat{\boldsymbol{\alpha}}_{.i} + \widehat{b}_i)} \text{ for } j \in R_i \text{ or } j \in S_i. \quad (3.5)$$

On the other hand, GWKLR is characterized by the penalty and kernel parameters. To choose optimal values of these hyperparameters of GWKLR we define a leave-one-out cross validation (LOOCV) function as follows;

$$CV(\boldsymbol{\theta}) = \frac{1}{\sum_{i=1}^{m} n_i} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - \widehat{p}_{ij}^{(-ij)})^2 \tag{3.6}$$

where $\boldsymbol{\theta}$ is a set of hyperparameters, $\widehat{p}_{ij}^{(-ij)} = \frac{\exp(K(\boldsymbol{x}_{ij}, \boldsymbol{x}^{(-ij)})\widehat{\boldsymbol{\alpha}}_{.i}^{(-ij)} + \widehat{b}_i^{(-ij)})}{1 + \exp(K(\boldsymbol{x}_{ij}, \boldsymbol{x}^{(-ij)})\widehat{\boldsymbol{\alpha}}_{.i}^{(-ij)} + \widehat{b}_i^{(-ij)})}$, $\boldsymbol{x}^{(-ij)} = \boldsymbol{x} \backslash \boldsymbol{x}_{ij}$ and $(\boldsymbol{\alpha}_{.i}^{(-ij)}, b_i^{(-ij)})$ is the parameter vector estimated without $(y_{ij}, \boldsymbol{x}_{ij})$ for $i \in S_i$.

4. Numerical studies

We now investigate the estimation performance of GWKLR through an artificial data set and the 1996 presidential election data set obtained from http://www. spatial-econometrics. com/html/jplv7.zip. We compare GWKLR with MELR and MEKLR. By the way, the location variable is treated as the input variable for MELR and MEKLR. We employ Gaussian kernel $K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp(-||\boldsymbol{x}_1 - \boldsymbol{x}_2||^2/\gamma)$, $\gamma > 0$, and the LOOCV function for model selection.

Example 4.1 We generate an artificial data set consisting of 10 small areas as follows:

1) Generate the location of center of each area, $\boldsymbol{u}_i$, from $U(-1,1)\times U(-1,1)$ for $i = 1, \cdots, 10$.
2) Determine N_i and n_i using $N_i = [100u_i]$ and $n_i = 0.1[N_i]$ for $i = 1, \cdots, 10$, where u_i is generated from $U(0.5, 1.5)$ and [] is round-off function.
3) Generate input variable x_{ij} from $U(0,1)$, which is nonlinearly related to the canonical parameter $\eta_{ij} = \sin(\pi x_{ij}) + \cos(\pi i) + l_{i1}l_{i2}$, where l_{ik} is the kth coordinate of the center of the ith area for $k = 1, 2$, and $j = 1, \cdots, N_i$ is the unit number in the ith area for $i = 1, \cdots, 10$. Generate the probability p_{ij} such as $p_{ij} = \exp(\eta_{ij})/(1 + \exp(\eta_{ij}))$ and the response y_{ij} from Bernoulli distribution $Ber(p(i))$ with $p(i) = \sum_{j=1}^{N_i} p_{ij}/N_i$.

To investigate the performance of GWKLR for the estimation of proportion of 10 areas, we use data sets randomly sampled from each area for training. This process is repeated 100 times ($t = 1, \cdots, 100$) to obtain the mean absolute error (MAE) and the mean-squared error (MSE) as follows;

$$MAE(i) = \frac{1}{100} \sum_{t=1}^{100} |r(i,t)|, \; MSE(i) = \frac{1}{100} \sum_{t=1}^{100} r(i,t)^2$$

where $r(i,t) = \widehat{p}(i,t) - p(i)$, $p(i)$ is the proportion of the ith area and $\widehat{p}(i,t)$ is the estimated proportion of the ith small area for $i = 1, \cdots, 10, t = 1, \cdots, 100$. Here $\widehat{p}(i,t)$ is the bias-adjusted estimate such as

$$\widehat{p}(i,t) = \frac{1}{N_i} \left(y_{tr}(i,t) + \widehat{y}_{ts}(i,t) + \frac{N_i - n_i}{n_i} (y_{tr}(i,t) - \widehat{y}_{tr}(i,t)) \right),$$

where $y_{tr}(i,t)$ is the number of units of favor among n_i sampled units from the ith area, $\widehat{y}_{tr}(i,t) = \sum_{j\in S(i,t)}[\widehat{p}(i,j,t)]$, $\widehat{y}_{ts}(i,t) = \sum_{j\in R(i,t)}[\widehat{p}(i,j,t)]$, $\widehat{p}(i,j,t)$is the estimate of probability of belongs to a certain class in the jth unit among n_i sampled units from the ith area at the tth iteration, $S(i,t)$ is the set of sampled units in the ith area at the tth iteration, $R(i,t)$ is the set of unsampled units in the ith area at the tth iteration. The numbers N_i and n_i for training are shown in Table 4.1.

Tables 4.1 and 4.2 show the performance results for MELR, MEKLR and GWKLR. As seen from tables, GWKLR generally works better than the others.

Table 4.1 Averages of estimates of proportion for each area in the artificial data

area	N_i	n_i	$p(i)$	$\widehat{p}(i)$ (MELR)	$\widehat{p}(i)$ (MEKLR)	$\widehat{p}(i)$ (GWKLR)
1	145	15	0.4951	0.4274	0.3998	0.4293
2	73	7	0.7892	0.7597	0.7079	0.6942
3	111	11	0.3967	0.3342	0.3300	0.3890
4	99	10	0.8119	0.7778	0.7558	0.7540
5	139	14	0.5576	0.5256	0.5184	0.5050
6	126	13	0.9091	0.8113	0.8103	0.8107
7	96	10	0.5232	0.4736	0.4759	0.5008
8	52	5	0.8410	0.7509	0.7418	0.7455
9	132	13	0.4271	0.3012	0.3012	0.3734
10	94	9	0.8412	0.7611	0.7512	0.7539

Table 4.2 Averages of 100 MAEs and 100 MSEs for estimates of p in artificial data (standard error in parenthesis)

state	MELR		MEKLR		GWKLR	
	MAE	MSE	MAE	MSE	MAE	MSE
1	0.1178 (0.0074)	0.0193 (0.0021)	0.1342 (0.0106)	0.0291 (0.0043)	**0.0941** (0.0066)	**0.0132** (0.0015)
2	0.1580 (0.0112)	0.0374 (0.0048)	0.1405 (0.0129)	0.0363 (0.0069)	**0.1363** (0.0107)	**0.0299** (0.0038)
3	0.1386 (0.0108)	0.0308 (0.0050)	0.1231 (0.0099)	0.0249 (0.0039)	**0.1071** (0.0085)	**0.0186** (0.0027)
4	0.0978 (0.0088)	0.0172 (0.0033)	0.0937 (0.0084)	0.0157 (0.0027)	**0.0893** (0.0073)	**0.0132(0.0019)**
5	0.1353 (0.0099)	0.0281 (0.0041)	**0.1044** (0.0086)	0.0183 (0.0027)	0.1058 (0.0084)	**0.0182** (0.0025)
6	0.0998 (0.0071)	0.0150 (0.0017)	0.0988 (0.0070)	0.0146 (0.0016)	**0.0985** (0.0069)	**0.0145** (0.0016)
7	0.1161 (0.0098)	0.0229 (0.0042)	0.1123 (0.0093)	0.0212 (0.0034)	**0.0992** (0.0086)	**0.0174** (0.0029)
8	**0.1442** (0.0112)	0.0333 (0.0061)	0.1482 (0.0110)	0.0340 (0.0057)	0.1445 (0.0110)	**0.0309** (0.0044)
9	0.1457 (0.0105)	0.0321 (0.0039)	0.1457 (0.0102)	0.0315 (0.0036)	**0.1016** (0.0072)	**0.0155** (0.0020)
10	0.1311 (0.0101)	0.0272 (0.0041)	0.1191 (0.0088)	0.0218 (0.0030)	**0.1176** (0.0088)	**0.0215** (0.0035)

Example 4.2 The 1996 presidential election data contain approval indices of Clinton for presidential election in 853 counties of 13 states. The location of each state (area) is given as the longitude and latitude of the capital of each state. The input variables are described in Table 4.3.

To show the performance of GWKLR for the estimation of approval ratings (proportion) for 13 states, for training we use data sets randomly sampled from each state. This process is repeated 100 times ($t = 1, \cdots, 100$) to obtain MAE and MSE. We use the bias-adjusted estimate (Salvati *et al.*, 2011), $\widehat{p}(i,t)$, which is computed as follows:

$$\widehat{p}(i,t) = \frac{1}{N_i}\left(y_{tr}(i,t) + \widehat{y}_{ts}(i,t) + \frac{N_i - n_i}{n_i}(y_{tr}(i,t) - \widehat{y}_{tr}(i,t)) \right),$$

where $y_{tr}(i,t)$ is the number of counties (units) of favor in Clinton among n_i sampled counties, $\widehat{y}_{tr}(i,t) = \sum_{j\in S(i,t)}[\widehat{p}(i,j,t)]$, $\widehat{y}_{ts}(i,t) = \sum_{j\in R(i,t)}[\widehat{p}(i,j,t)]$, $\widehat{p}(i,j,t)$is the estimate of approval rating of jth county among n_i sampled counties at the tth iteration, $S(i,t)$ is the set of sampled counties in the ith state at the tth iteration, $R(i,t)$ is the set of unsampled counties in the ith state at the tth iteration. N_i and n_i are shown in Table 4.4.

Tables 4.4 and 4.5 show the performance results for MELR, MEKLR and GWKLR. In Table 4.5 the boldfaced figure in each row signifies the smallest value of MAE and MSE for a given state. As seen from tables, GWKLR generally works better than the others.

Table 4.3 Description of input variables for election data

Variables	Description
x_1	Logarithm of urban population
x_2	Logarithm of rural population
x_3	Population with high school or GED graduates as a proportion of educated
x_4	Population with some college as a proportion of educated
x_5	Population with associate degrees as a proportion of educated
x_6	Population with college degrees as a proportion of educated
x_7	Population with grad/professional degrees as a proportion of educated

Table 4.4 Estimates of the approving rating of Clinton for each state in election data

state	N_i	n_i	$p(i)$	MELR	MEKLR	GWKLR
1	67	10	0.4627	0.3932	0.3467	0.4057
2	75	11	0.8800	0.7369	0.7798	0.7403
3	15	8	0.6000	0.1884	0.2676	0.2048
4	58	9	0.4310	0.3908	0.4064	0.3825
5	63	9	0.3492	0.3395	0.2924	0.3213
6	9	5	1.0	0.4444	0.4464	0.4444
7	3	2	1.0	0.3333	0.3350	0.3333
8	67	10	0.4776	0.4143	0.3797	0.3953
9	159	25	0.4906	0.4390	0.3603	0.4311
10	99	15	0.2020	0.2154	0.1244	0.2092
11	44	7	0.9091	0.7424	0.7634	0.7471
12	102	15	0.3725	0.3762	0.2455	0.3722
13	92	14	0.7174	0.5815	0.5826	0.5883

Table 4.5 Averages of 100 MAEs and 100 MSEs for estimates of the approving rating of Clinton in election data (standard error in parenthesis)

state	MEKLR		MEKLR		GWKLR	
	MAE	MSE	MAE	MSE	MAE	MSE
1	0.1456 (0.0126)	0.0369 (0.0072)	0.1235 (0.0095)	0.0222 (0.0038)	**0.1028** (0.0084)	**0.0181** (0.0027)
2	0.1244 (0.0106)	0.0267 (0.0043)	**0.0940** (0.0077)	**0.0158** (0.0024)	0.1403 (0.0097)	0.0290 (0.0035)
3	**0.3346** (0.0189)	**0.1473** (0.0132)	0.3400 (0.0217)	0.1616 (0.0161)	0.3946 (0.0122)	0.1722 (0.0107)
4	0.1259 (0.0084)	0.0228 (0.0028)	0.0959 (0.0083)	0.0155 (0.0025)	**0.0800** (0.0079)	**0.0125** (0.0024)
5	0.1565 (0.0110)	0.0365 (0.0044)	0.1235 (0.0093)	0.0232 (0.0034)	**0.0955** (0.0079)	**0.0161** (0.0025)
6	0.5573 (0.0046)	0.3127 (0.0048)	**0.5556** (0.0000)	**0.3086** (0.0000)	**0.5556** (0.0000)	**0.3086** (0.0000)
7	**0.6400** (0.0170)	**0.4383** (0.0187)	0.6667 (0.0077)	0.4461 (0.0105)	0.6667 (0.0000)	0.4444 (0.0000)
8	0.1453 (0.0113)	0.0339 (0.0048)	0.1148 (0.0084)	0.0204 (0.0025)	**0.1049** (0.0085)	**0.0181** (0.0026)
9	0.1067 (0.0086)	0.0186 (0.0026)	0.1117 (0.0074)	0.0170 (0.0022)	**0.0684** (0.0049)	**0.0072** (0.0011)
10	0.1021 (0.0074)	0.0158 (0.0023)	0.1096 (0.0064)	0.0172 (0.0014)	**0.0757** (0.0061)	**0.0094** (0.0012)
11	0.1791 (0.0136)	0.0503 (0.0081)	0.1644 (0.0130)	0.0434 (0.0074)	**0.1610** (0.0112)	**0.0383** (0.0052)
12	0.1267 (0.0089)	0.0239 (0.0030)	0.1346 (0.0080)	0.0246 (0.0025)	**0.0734** (0.0060)	**0.0086** (0.0015)
13	0.1482 (0.0107)	0.0333 (0.0040)	**0.1376** (0.0083)	**0.0260** (0.0027)	0.1419 (0.0085)	0.0273 (0.0029)

5. Conclusion

We investigated how well GWKLR works for SAE of the proportion. GWKLR actually takes over advantages of kernel machine and GWR that capture spatial information in data. We found that GWKLR is a efficient technique when the functional forms of the relationship between the response variable and the input variables are under the complex patterns of spatial dependence. We also found that GWKLR can be effectively employed without heavy computations when input variables are high dimensional.

References

Anselin, L. (1992). *Spatial econometrics: Method and models*, Kluwer Academic Publishers, Boston.

Brunsdon, C., Fotheringham, A. S. and Charlton, M. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, **28**, 281-298.

Harville, D. A. (1976). Extension of Gauss-Markov theorem to include the estimation of random effects. *Annals of Statistics*, **4**, 384-395.

Hwang, H. (2014). Support vector quantile regression for autoregressive data. *Journal of the Korean Data & Information Science Society*,**25**, 1539-1547.

Hwang, H. (2015). Partially linear support vector orthogonal quantile regression with measurement errors. *Journal of the Korean Data & Information Science Society*, **26**, 209-216.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, **15**, 1-96.

Kimeldorf, G. and Wahba, G. (1971). Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, **33**, 82-95.

Mercer, J. (1909). Functions of positive and negative type and their connection with theory of integral equations. *Philosophical Transactions of Royal Society A*, 415-446.

Perez-Cruz, F., Bousono-Calzon, C. and Artes-Rodriguez, A. (2005). Convergence of the IRWLS procedures to SVM solution. *Neural Computation*, **17**, 7-18.

Pfeffermann, D, (2013). New important developments in small area estimation. *Statistical Science*, **28**, 40-68.

Rao, J. N. K. (2003). *Small area estimation*, Wiley, New York.

Rifkin, R. and Klautau, A. (2004). In defense of one-vs-all classification. *Journal of Machine Learning Research*, **5**, 101-141.

Salvati, N., Ranalli, M. G. and Pratesi, M. (2011). Small area estimation of the mean using non-parametric M-quantile regression: A comparison when a linear mixed model does not hold. *Journal of Statistical Computation and Simulation*, **81**, 945- 964.

Seok, K. (2015). Semisupervised support vector quantile regression. *Journal of the Korean Data & Information Science Society*, **26**, 517-524.

Shim, J. and Hwang, C. (2014). Estimating small area proportions with kernel logistic regressions models. *Journal of the Korean Data & Information Science Society*, **25**, 941-949.

Shim, J. and Seok, K. H. (2014). A transductive least squares support vector machine with the difference convex algorithm. *Journal of the Korean Data & Information Science Society*, **25**, 455-464.

Suykens, J. A. K. and Vandewalle, J. (1999). Least squares support vector machine classiers. *Neural Processing Letters*, **9**, 293-300.

Vapnik, V. N. (1995). *The nature of statistical learning theory*, Springer, New York.

Vapnik, V. N. (1998). *Statistical learning theory*, Springer, New York.

Yu, D. L. and Wu, C. (2004) Understanding population segregation from Landsat ETM+imagery: A geographically weighted regression approach. *GISience and Remote Sensing*, **41**, 145-164.