

# Kernel Poisson regression for mixed input variables<sup>†</sup>

Jooyong Shim<sup>1</sup>

<sup>1</sup>Department of Data Science, Inje University

Received 17 October 2012, revised 31 October 2012, accepted 5 November 2012

## Abstract

An estimating procedure is introduced for kernel Poisson regression when the input variables consist of numerical and categorical variables, which is based on the penalized negative log-likelihood and the component-wise product of two different types of kernel functions. The proposed procedure provides the estimates of the mean function of the response variables, where the canonical parameter is linearly and/or nonlinearly related to the input variables. Experimental results are then presented which indicate the performance of the proposed kernel Poisson regression.

*Keywords:* Canonical parameter, kernel function, partially linear regression, penalized negative log-likelihood, Poisson regression.

## 1. Introduction

Poisson regression is widely used to analyze the event count data (Vermunt, 1996). It can be used to model the number of occurrences of an event of interest or the rate of occurrence of an event of interest, as a function of some input variables. A nonparametric estimation for the canonical parameter (logarithm of the mean function) of Poisson process based on penalized likelihood smoothing spline models was proposed by O'Sullivan *et al.* (1986). When canonical parameter is of a linear form, it is known that the generalized linear model with a Poisson likelihood (McCullagh and Nelder, 1983) is useful to Poisson process. But the linear assumption on the canonical parameter is strict in some cases. Various parametric approaches are proposed by Wei (1998) to attain more flexibility than a simple linear model. Kernel technique applied in support vector machine (Vapnik, 1998) have been very successful in nonlinear problems. Applications of kernel technique can be found in Cho *et al.* (2010) and Shim and Hwang (2010).

In this paper we propose a kernel Poisson regression to obtain the estimates of the mean function of response variables when the input vector is divided into two vectors - a vector of numerical variables and a vector of categorical variables. The rest of this paper is organized as follows. In Section 2 we give a kernel Poisson regression based on the penalized negative log-likelihood. In Section 3 we give a brief review of partially linear kernel Poisson regression. In Section 4 we propose the component-wise product of two kernel functions for kernel Poisson regression with mixed input variables. In Section 5 we perform the numerical studies through examples. In Section 6 we give the conclusions.

<sup>†</sup> 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 (2012000646).

<sup>1</sup> Adjunct professor, Department of Data Science, Institute of Statistical Information, Inje University, Kimhae 621-749, Korea. E-mail: ds1631@hanmail.net.

## 2. Kernel Poisson regression

In Poisson regression it is assumed that the response variable  $y_i \in \{0, 1, \dots\}$ , number of occurrences of an event given the input vector  $\mathbf{x}_i \in \mathbf{R}^d$ , follows a Poisson distribution,

$$p(y_i|\mathbf{x}_i) = \frac{\exp(-\mu(\mathbf{x}_i))\mu(\mathbf{x}_i)^{y_i}}{y_i!}, \quad i = 1, 2, \dots, n. \quad (2.1)$$

The negative log-likelihood of the given data set can be expressed as (a constant term is omitted)

$$\ell(\boldsymbol{\mu}) = \frac{1}{n} \sum_{i=1}^n (\mu(\mathbf{x}_i) - y_i \log \mu(\mathbf{x}_i)). \quad (2.2)$$

We write the canonical parameter (logarithm of  $\mu(\mathbf{x}_i)$ ) as  $\eta(\mathbf{x}_i)$ , then the negative log-likelihood can be reexpressed as

$$\ell(\boldsymbol{\eta}) = \frac{1}{n} \sum_{i=1}^n (\exp(\eta(\mathbf{x}_i)) - y_i \eta(\mathbf{x}_i)). \quad (2.3)$$

The canonical parameter given  $\mathbf{x}_i$  is estimated by a linear model,  $\eta(\mathbf{x}_i) = \boldsymbol{\omega}'\phi(\mathbf{x}_i) + b$ , conducted in a high dimensional feature space. Here the feature mapping function  $\phi(\cdot) : \mathbf{R}^d \rightarrow \mathbf{R}^{d_f}$  maps the input space to the higher dimensional feature space where the dimension  $d_f$  is defined in an implicit way. It is known that  $\phi(\mathbf{x}_i)'\phi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$  which are obtained from the application of Mercer's conditions (1909). Then the estimate of canonical parameter vector  $\boldsymbol{\eta}(\mathbf{x}) = \boldsymbol{\omega}'\phi(\mathbf{x}) + b$  for  $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^n$  is obtained by minimizing the penalized negative log-likelihood,

$$\ell(\boldsymbol{\omega}, b) = \frac{1}{n} \sum_{i=1}^n (\exp(\boldsymbol{\omega}'\phi(\mathbf{x}_i) + b) - y_i(\boldsymbol{\omega}'\phi(\mathbf{x}_i) + b)) + \frac{\lambda}{2} \|\boldsymbol{\omega}\|^2 \quad (2.4)$$

where  $\lambda$  is a nonnegative constant which controls the trade-off between the goodness-of-fit on the data and  $\|\boldsymbol{\omega}\|^2$ . The representation theorem (Kimeldorf and Wahba, 1971) guarantees the minimizer of the penalized negative log-likelihood to be  $\boldsymbol{\eta}(\mathbf{x}) = K(\mathbf{x}, \mathbf{x})\boldsymbol{\alpha} + b$  for some  $n \times 1$  vector  $\boldsymbol{\alpha}$  and a scalar  $b$ , where  $K(\mathbf{x}, \mathbf{x})$  is the  $n \times n$  kernel function constructed with  $\mathbf{x}$ .

Now the penalized negative log-likelihood (2.4) becomes

$$\ell(\boldsymbol{\alpha}, b) = \frac{1}{n} \sum_{i=1}^n (\exp(K(\mathbf{x}_i, \mathbf{x})\boldsymbol{\alpha} + b) - y_i(K(\mathbf{x}_i, \mathbf{x})\boldsymbol{\alpha} + b)) + \frac{\lambda}{2} \boldsymbol{\alpha}'K(\mathbf{x}, \mathbf{x})\boldsymbol{\alpha}. \quad (2.5)$$

By minimizing the penalized negative log-likelihood (2.5) we obtain the estimator of parameter vector  $\boldsymbol{\alpha}$  and  $b$ , but not in an explicit form, which leads to use the iterative method. At each step the parameter vector  $\boldsymbol{\alpha}$  and  $b$  are updated by Newton-Raphson method as follows:

$$\begin{pmatrix} \boldsymbol{\alpha}^{(t+1)} \\ b^{(t+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}^{(t)} \\ b^{(t)} \end{pmatrix} - \begin{pmatrix} KD^{(t)}K + \lambda K & K\boldsymbol{\mu}^{(t)} \\ \boldsymbol{\mu}^{(t)'}K & \mathbf{1}'\boldsymbol{\mu}^{(t)} \end{pmatrix}^{-1} \begin{pmatrix} K(\boldsymbol{\mu}^{(t)} - \mathbf{y} + \lambda\boldsymbol{\alpha}^{(t)}) \\ \mathbf{1}'(\boldsymbol{\mu}^{(t)} - \mathbf{y}) \end{pmatrix}, \quad (2.6)$$

where  $\boldsymbol{\mu}^{(t)} = \exp(K\boldsymbol{\alpha}^{(t)} + b^{(t)})$  and  $D^{(t)}$  is a diagonal matrix of  $\boldsymbol{\mu}^{(t)}$ .

With the optimal values of  $\boldsymbol{\alpha}$  and  $b$ , the predicted mean function given the input vector  $\mathbf{x}_0$  is obtained as follows:

$$\hat{\mu}(\mathbf{x}_0) = \exp(K(\mathbf{x}_0, \mathbf{x})\boldsymbol{\alpha} + b) \quad (2.7)$$

### 3. Partially linear kernel Poisson regression

When the canonical parameter  $\eta_i$  is nonlinearly related to the input vector  $\mathbf{x}_i$  and linearly related to  $\mathbf{z}_i$  we consider the partially linear model of the form,

$$\eta_i = \boldsymbol{\omega}'\phi(\mathbf{x}_i) + \boldsymbol{\beta}'\mathbf{z}_i + b = K(\mathbf{x}_i, \mathbf{x})\boldsymbol{\alpha} + \boldsymbol{\beta}'\mathbf{z}_i + b. \tag{3.1}$$

Then the penalized negative log-likelihood can be written as

$$\ell(\boldsymbol{\alpha}, b) = \frac{1}{n} \sum_{i=1}^n (\exp(K(\mathbf{x}_i, \mathbf{x})\boldsymbol{\alpha} + \boldsymbol{\beta}'\mathbf{z}_i + b) - y_i(K(\mathbf{x}_i, \mathbf{x})\boldsymbol{\alpha} + \boldsymbol{\beta}'\mathbf{z}_i + b)) + \frac{\lambda}{2}\boldsymbol{\alpha}'K(\mathbf{x}, \mathbf{x})\boldsymbol{\alpha}. \tag{3.2}$$

By minimizing the penalized negative log-likelihood (3.2) we obtain the estimator of parameter vector  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$  and  $b$ , but not in a explicit form, which leads to use the iterative method. At each step the parameter vector  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$  and  $b$  are updated by Newton-Raphson method as follows:

$$\begin{pmatrix} \boldsymbol{\alpha}^{(t+1)} \\ \boldsymbol{\beta}^{(t+1)} \\ b^{(t+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}^{(t)} \\ \boldsymbol{\beta}^{(t)} \\ b^{(t)} \end{pmatrix} - \begin{pmatrix} KD^{(t)}K + \lambda K & KD^{(t)}\mathbf{z}' & K\boldsymbol{\mu}^{(t)} \\ \mathbf{z}D^{(t)}K & \mathbf{z}'D^{(t)}\mathbf{z} & \mathbf{z}'\boldsymbol{\mu}^{(t)} \\ \boldsymbol{\mu}^{(t)'}K & \boldsymbol{\mu}^{(t)'}\mathbf{z} & \mathbf{1}'\boldsymbol{\mu}^{(t)} \end{pmatrix}^{-1} \begin{pmatrix} K(\boldsymbol{\mu}^{(t)} - \mathbf{y} + \lambda\boldsymbol{\alpha}^{(t)}) \\ \mathbf{z}'(\boldsymbol{\mu}^{(t)} - \mathbf{y}) \\ \mathbf{1}'(\boldsymbol{\mu}^{(t)} - \mathbf{y}) \end{pmatrix}, \tag{3.3}$$

where  $\mathbf{z} = \{\mathbf{z}_i\}_{i=1}^n$ ,  $\boldsymbol{\mu}^{(t)} = \exp(K\boldsymbol{\alpha}^{(t)} + \mathbf{z}\boldsymbol{\beta}^{(t)} + b^{(t)})$  and  $D^{(t)}$  is a diagonal matrix of  $\boldsymbol{\mu}^{(t)}$ . With the optimal values of  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$  and  $b$ , the predicted mean function given the input vector  $(\mathbf{x}_0, \mathbf{z}_0)$  is obtained as follows:

$$\hat{\mu}(\mathbf{x}_0, \mathbf{z}_0) = \exp(K(\mathbf{x}_0, \mathbf{x})\boldsymbol{\alpha} + \mathbf{z}_0\boldsymbol{\beta} + b). \tag{3.4}$$

### 4. Kernel Poisson regression with mixed input variables

We assume that the input vector  $\mathbf{x}_i$  is divided into two vectors of  $\mathbf{x}_i^{(n)}$  and  $\mathbf{x}_i^{(c)}$ , where  $\mathbf{x}_i^{(n)}$  is a vector of numerical variables and  $\mathbf{x}_i^{(c)}$  is a vector of categorical variables. In this case we can apply the partially linear kernel Poisson regression in (3.1) by setting  $z_{ik} = 1$  if the  $i$ th observation belongs to the  $k$ th category, 0 else.

In this section we propose the simple way of estimating the mean function for the above case. We define kernel function of input variables by using the weighted exponential kernel function and Hamming distance kernel function (Aradhya and Dorai, 2002) as follows:

$$\begin{aligned} K(\mathbf{x}_i, \mathbf{x}_j) &= K^{(n)}(\mathbf{x}_i^{(n)}, \mathbf{x}_j^{(n)})K^{(c)}(\mathbf{x}_i^{(c)}, \mathbf{x}_j^{(c)}) \\ &= \exp\left(-\frac{1}{\sigma^2} \sum_{k=1}^{d_n} \frac{|x_{ik}^{(n)} - x_{jk}^{(n)}|}{d_n R_k} - \frac{1}{\beta} \sum_{l=1}^{d_c} \frac{I(x_{il}^{(c)} \neq x_{jl}^{(c)})}{d_c}\right) \end{aligned} \tag{4.1}$$

where  $d_n$  and  $d_c$  are the numbers of numerical variables and categorical variables, respectively.  $R_k$  is the range of the  $k$ -th numerical input variable,  $\sigma^2$  and  $\beta$  are kernel parameters. This kernel function can be expressed as a component-wise product of the weighted exponential kernel function and Hamming distance kernel function (Aradhya and Dorai, 2002), which leads an easy verification of kernelness (Genton, 2001).

By minimizing the penalized negative log-likelihood (2.5) we obtain the estimator of parameter vector  $\alpha$  and  $b$ , which leads that the predicted mean function given the input vector  $\mathbf{x}_t = (\mathbf{x}_t^{(n)}, \mathbf{x}_t^{(c)})$  is obtained as follows:

$$\hat{\mu}(\mathbf{x}_t, \mathbf{z}_t) = \exp(K(\mathbf{x}_t, \mathbf{x})\alpha + b). \tag{4.2}$$

The kernel function above is valid for missing data problems by introducing the weight  $\omega_{ijk}$ ,

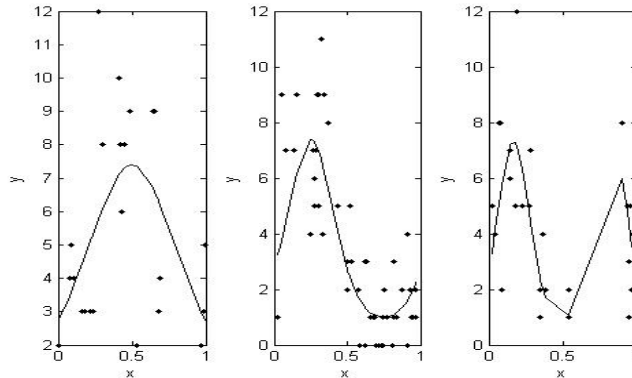
$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{1}{\sigma^2} \sum_{k=1}^{d_n} \frac{\omega_{ijk} |x_{ik}^{(n)} - x_{jk}^{(n)}|}{d_n R_k} - \frac{1}{\beta} \sum_{l=1}^{d_c} \frac{\omega_{ijl} \mathbf{I}(x_{il}^{(c)} \neq x_{jl}^{(c)})}{d_c}\right), \tag{4.3}$$

which is set to 0 if the value of  $k$ th input variable is missing for one or both observations, else 1.

### 5. Numerical studies

We illustrate the performance of the proposed kernel Poisson regression (the kernel Poisson regression with mixed input variables) through the simulated data and real data.

**Example 5.1** 100 training data sets of size 100 and 100 test data sets of size 50 are generated to present the estimation performance of the proposed procedure. Each data set consists of  $x$ 's,  $z$ 's and  $y$ 's. Here  $x$ 's are randomly generated from a uniform distribution  $(0, 1)$ ,  $z$ 's are randomly generated from a discrete uniform distribution  $(1, 2, 3)$  and  $y$ 's are generated from a Poisson distribution with the canonical parameter  $\eta(x_i, z_i) = \sin(\pi z_i x_i) + 1$ . Figure 5.1 shows true mean functions (solid line) imposed on the scatter plots of one of 100 data sets according to values of  $z_i$  (1,2,3).



**Figure 5.1** True mean functions (solid line) imposed on the scatter plots according to values of  $z_i$  (1,2,3)

For the Poisson regression the canonical parameter given  $(x_i, z_i)$  is modelled as  $\eta(x_i, z_i) = x_i \alpha + \mathbf{v}_i \beta + b$ , where  $\mathbf{v}_i = (1, 0, 0)$  if  $z_i = 1$ ,  $\mathbf{v}_i = (0, 1, 0)$  if  $z_i = 2$  and  $\mathbf{v}_i = (0, 0, 1)$  if  $z_i = 3$ . For the partially linear kernel Poisson regression the canonical parameter given  $(x_i, z_i)$  is

modelled as  $\eta(x_i, z_i) = K(x_i, \mathbf{x})\boldsymbol{\alpha} + \mathbf{v}_i\boldsymbol{\beta} + b$ . The radial basis kernel function is utilized for the partially linear kernel Poisson regression, which is

$$K(x_k, x_l) = \exp\left(-\frac{1}{\sigma^2}\|x_k - x_l\|^2\right).$$

For the proposed kernel Poisson regression the canonical parameter given  $(x_i, z_i)$  is modelled as  $\eta(x_i, z_i) = K((x_i, z_i), (\mathbf{x}, \mathbf{z}))\boldsymbol{\alpha} + b$ , where  $K((x_i, z_i), (x_j, z_j)) = \exp(-(1/\sigma^2)\|x_i - x_j\|/R - (1/\beta)I(z_i \neq z_j))$ .

For the model selection of the partially linear kernel Poisson regression and the kernel Poisson regression with mixed input variables, we use 5-fold cross validation function. To illustrate the estimation performance of the proposed kernel Poisson regression, the mean squared error (mse) and residual deviance (dev) are used as the prediction performance measure defined by

$$mse_k = \frac{1}{n} \sum_{i=1}^n (\mu_i - \hat{\mu}_i^{(k)})^2 \text{ and } dev_k = 2 \sum_{i=1}^n \left( y_i \log \left( y_i / \hat{\mu}_i^{(k)} \right) - (y_i - \hat{\mu}_i^{(k)}) \right), \quad k = 1, 2, 3,$$

where  $\mu_i$  is the true mean function,  $\hat{\mu}_i^{(1)}$  is the estimated mean function obtained by Poisson regression,  $\hat{\mu}_i^{(2)}$  is the estimated mean function obtained by the partially linear kernel Poisson regression and  $\hat{\mu}_i^{(3)}$  is the estimated mean function obtained by the proposed kernel Poisson regression. From 100 training data sets and 100 test data sets we obtained the means, standard deviations and medians of  $mse_k$ 's and  $dev_k$ 's for  $k = 1, 2, 3$  as in Table 5.1 and Table 5.2. Box plots of  $mse_k$ 's and  $dev_k$ 's for 100 training data sets and 100 test data sets are shown in Figure 5.2 and Figure 5.3. In the figures and tables we can see that the proposed kernel Poisson regression provide better prediction performance than other two Poisson regressions in this example.

**Table 5.1** Means, standard deviations and medians of  $mse_k$ 's and  $dev_k$ 's for 100 training data sets

	$mse_1$	$mse_2$	$mse_3$	$dev_1$	$dev_2$	$dev_3$
mean	4.2715	3.5523	1.0301	217.6786	197.0582	79.9693
std dev	0.4026	0.5567	0.5018	26.0854	26.5112	21.7871
median	4.3346	3.5775	0.8901	214.5170	194.1346	86.4980

**Table 5.2** Means, standard deviations and medians of  $mse_k$ 's and  $dev_k$ 's for 100 test data sets

	$mse_1$	$mse_2$	$mse_3$	$dev_1$	$dev_2$	$dev_3$
mean	4.4176	3.8668	1.2470	113.1261	107.5929	69.3982
std dev	0.5987	1.2859	0.6391	20.0747	20.3108	14.6333
median	4.4475	3.7773	1.0737	111.9272	107.1997	67.4719

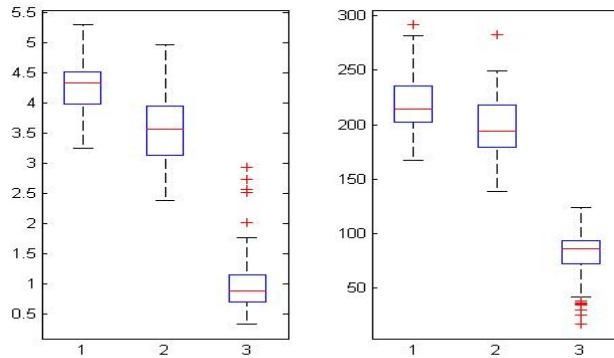


Figure 5.2 Box plots of  $mse'_k$ 's (left) and  $dev'_k$ 's (right) for 100 training data sets

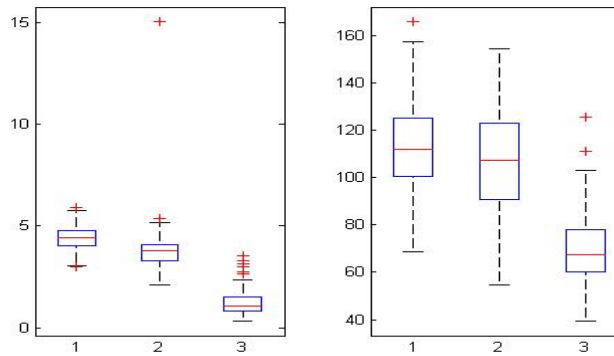


Figure 5.3 Box plots of  $mse'_k$ 's (left) and  $dev'_k$ 's (right) for 100 test data sets

**Example 5.2** We use Swedish third-party motor insurance data of the cities of Stockholm, Göteborg and Malmo in 1977 (Hallin and Ingenbleek, 1983). The 295 data consist of kilometers travelled per year ( $x_1$ , 1 : < 1000, 2: 1000 – 15000, 3: 15000 – 20000, 4: 20000 – 25000, 5: > 25000), the number of policy years plus one since last claim ( $x_2$ ), car make ( $z$ , 1-8: represent eight different common car models) and number of claims ( $y$ ). Figure 5.4 shows the scatter plots of number of claims versus kilometers travelled per year according to car make (1,  $\dots$ , 8). We randomly divide the data set into the training data set of size 195 and the test data set of size 100. We repeated this procedure 50 times.

For the Poisson regression the canonical parameter given  $(\mathbf{x}_i, z_i)$  is modelled as  $\eta(\mathbf{x}_i, z_i) = \mathbf{x}_i\boldsymbol{\alpha} + \mathbf{v}_i\boldsymbol{\beta} + b$ , where  $\mathbf{v}_i = (1, 0, 0, 0, 0, 0, 0, 0)$  if  $z_i = 1, \dots, \dots, \mathbf{v}_i = (0, 0, 0, 0, 0, 0, 0, 1)$  if  $z_i = 8$ . For the partially linear kernel Poisson regression the canonical parameter given  $(\mathbf{x}_i, z_i)$  is modelled as  $\eta(\mathbf{x}_i, z_i) = K(\mathbf{x}_i, \mathbf{x})\boldsymbol{\alpha} + \mathbf{v}_i\boldsymbol{\beta} + b$ . The radial basis kernel function is utilized in the partially linear kernel Poisson regression, which is

$$K(\mathbf{x}_k, \mathbf{x}_l) = \exp\left(-\frac{1}{\sigma^2}\|\mathbf{x}_k - \mathbf{x}_l\|^2\right).$$

For the proposed kernel Poisson regression (the kernel Poisson regression with mixed input

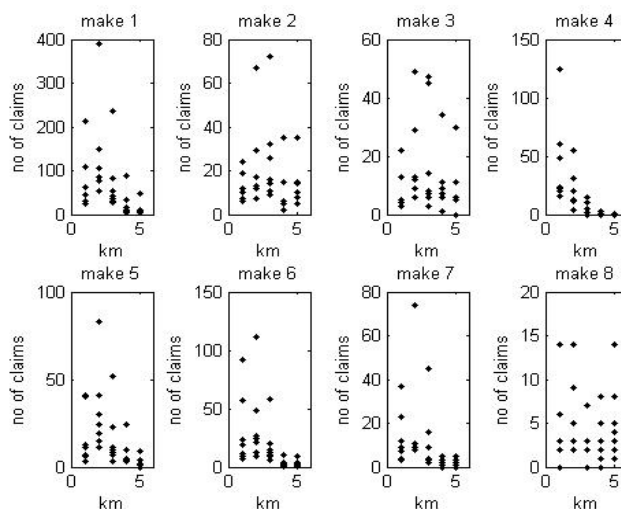
variables) the canonical parameter given  $(\mathbf{x}_i, z_i)$  is modelled as  $\eta(\mathbf{x}_i, z_i) = K((\mathbf{x}_i, z_i), (\mathbf{x}, \mathbf{z}))\boldsymbol{\alpha} + b$ , where  $K((\mathbf{x}_i, z_i), (\mathbf{x}_j, z_j)) = \exp(-1/2\sigma^2 \sum_{k=1}^2 |x_{ik} - x_{jk}|/R_k - (1/\beta)I(z_i \neq z_j))$ .

For the model selection of the partially linear kernel Poisson regression and the proposed kernel Poisson regression, we use 5-fold cross validation function.

To illustrate the estimation performance of the proposed kernel Poisson regression, the mean squared error (mse) and residual deviance (dev) are used as the prediction performance measure defined by

$$mse_k = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}_i^{(k)})^2 \text{ and } dev_k = 2 \sum_{i=1}^n \left( y_i \log \left( y_i / \hat{\mu}_i^{(k)} \right) - (y_i - \hat{\mu}_i^{(k)}) \right), k = 1, 2, 3.$$

From 50 training data sets and 50 test data sets we obtained the means, standard deviations and medians of  $mse_k$ 's and  $dev_k$ 's for  $k = 1, 2, 3$  as in Table 5.3 and Table 5.4. Box plots of  $mse_k$ 's and  $dev_k$ 's for 50 training data sets and 50 test data sets are shown in Figure 5.5 and Figure 5.6. In the figures and tables we can see that the proposed kernel Poisson regression provide better prediction performance than other two Poisson regressions in the given motor insurance data.



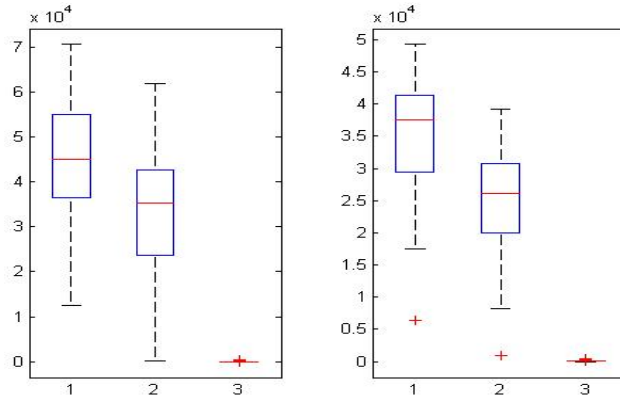
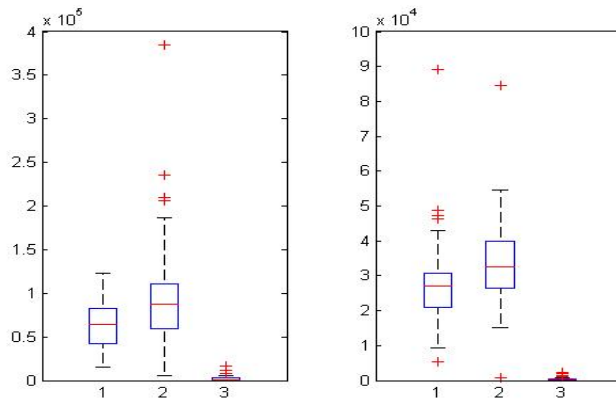
**Figure 5.4** Scatter plots of number of claims versus kilometers travelled per year according to car make (1, . . . , 8) for Swedish third-party motor insurance data

**Table 5.3** Means, standard deviations and medians of  $mse_k$ 's and  $dev_k$ 's for 50 training data sets

	$mse_1$	$mse_2$	$mse_3$	$dev_1$	$dev_2$	$dev_3$
mean	45298	33172	16.212	35016	24997	94.800
std dev	13823	14646	46.761	9145.3	7840.0	100.60
median	45029	35354	4.0358	37576	26087	73.704

**Table 5.4** Means, standard deviations and medians of  $mse_k$ 's and  $dev_k$ 's for 50 test data sets

	$mse_1 \times 10^{-6}$	$mse_2 \times 10^{-5}$	$mse_3 \times 10^{-3}$	$dev_1 \times 10^{-4}$	$dev_2 \times 10^{-4}$	$dev_3$
mean	4.1238	1.9026	2.2780	3.5169	3.3165	488.70
std dev	27.631	6.7034	2.9707	5.2267	1.2113	393.30
median	0.0646	0.8716	1.1180	2.7148	3.2700	361.01

**Figure 5.5** Box plots of  $mse'_k$ 's (left) and  $dev'_k$ 's (right) for 50 training data sets**Figure 5.6** Box plots of  $mse'_k$ 's (left) and  $dev'_k$ 's (right) for 50 test data sets

## 6. Conclusions

In this paper, we dealt with estimating the mean function by the kernel Poisson regression when input variables consist of numerical and categorical variables. Through the simulated data and real data we showed that the proposed kernel Poisson regression derives more satisfying results than Poisson regression and the partially linear kernel Poisson regression.



## References

- Aradhye, H. and Dorai, C. (2002). New kernels for analyzing multimodal data in multimedia using kernel machines. *Proceedings of the IEEE International Conference on Multimedia and Expo*, **2**, 37-40.
- Cho, D. H., Shim, J. and Seok, K. H. (2010). Doubly penalized kernel method for heteroscedastic autoregressive data. *Journal of the Korean Data & Information Science Society*, **21**, 155-162.
- Genton, M. G. (2001). Classes of kernels for machine learning: A statistics perspective. *Journal of Machine Learning Research*, **2**, 299-312.
- Hallin, M. and Ingenbleek, J. F. (1983). The Swedish automobile portfolio in 1977. A statistical study. *Scandinavian Actuarial Journal*, 49-64.
- Kimeldorf, G. S. and Wahba, G. (1971). Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and its Applications*, **33**, 82-95.
- McCullagh, P. and Nelder, J. A. (1983). *Generalized linear models*, Chapman & Hall, London.
- 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.
- O'Sullivan, F., Yandell, B. S. and Raynor, W. J. (1986). Automatic smoothing of regression functions in generalized linear models. *Journal of American Statistical Associations*, **81**, 96-103.
- Shim, J. and Hwang, C. (2010). Semiparametric support vector machine for accelerated failure time model. *Journal of the Korean Data & Information Science Society*, **21**, 467-477.
- Vapnik, V. N. (1998). *Statistical learning theory*, Wiley, New York.
- Vermunt, J. K. (1996). *Log-linear event history analysis*, Series on work and organization, Tilburg University Press, Tilburg.
- Wei, B. C. (1998). Exponential family nonlinear models. *Lecture Notes in Statistics*, **130**, Springer-Verlag Singapore, Singapore.