International Journal of Advances in Computer Science & Its Applications– IJCSIA
Volume 5: Issue 1 [ISSN : 2250-3765]
Publication Date : 30 April, 2015

Spline Kernels in Nonparametric Regression

Marcin Michalak

***Abstract*—This paper presents a new group of kernel functions – spline kernels. This group of kernels connects advantages of Epanechnikov kernel and a multiple differentiability, required in several aspects of building kernel estimators.**

***Keywords*—nonparametric regression, kernel estimators, kernel function**

I. Introduction

Estimation of the regression function is a very common problem in machine learning [14][16]. Its aim is to find the hidden dependencies between known and modelled variables.

These methods are divided into two groups: parametric and nonparametric. In the first group we are given a data and a class of function, in which we will be searching for the best fit into the data. In other word – the pattern with a finite set of parameters is given (e.g. the linear function requires setting or estimation of two parameters) and the aim of parametric regression is to find the best values of these parameters. In a case of nonparametric regression no pattern is given.

The most popular methods of nonparametric regression are spline functions [2][5], additive (and generalized additive) models [8], LOWESS algorithm 0, Support Vector Machines [13] and kernel estimators [11][18].

In this paper the new model of local kernel algorithms (with the limited domain where it takes non-zero values) is presented. Next section presents the backgrounds of nonparametric kernel regression estimator with a special attention paid to the method of selection of smoothing parameter. Then the definition and construction of spline kernels is described. It is followed by the part with the analysis of statistical properties of newly created kernels. The paper ends with a short discussion.

II. Nonparametric Regression Function Estimation

Nonparametric methods of regression function estimation can be described as a kind of black-box models. It is due to the fact, that we are given an answer of the model for a given input, but the nature of giving this answer cannot be interpreted. In this section kernel estimators – a very popular branch of nonparametric methods – are presented with a short overview of methods of smoothing parameter calculation.

Marcin Michalak
Institute of Informatics, Silesian University of Technology
ul. Akademicka 16
44-100 Gliwice, Poland

A. *Kernel Estimators*

Kernel estimator is generally a mapping function $\tilde{f}(x): R^n \to R$ of the following form:

$$\tilde{f}(x) = \frac{1}{nh^m} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

where n is the number of training objects and h is an operator called a smoothing operator. Most popular kernel estimators are Nadaraya-Watson [11][18], Stone-Fan 0, Priestley-Chao [17] and Gasser-Mueller 0.

Kernel function must fulfils the following conditions:

- $\int_R K(x)dx = 1$
- $\int_R xK(x)dx = 0$
- $\int_R x^2K(x)dx < \infty$
- $\forall x \in R\ K(0) \geq K(x)$
- $\forall x \in R\ K(x) = K(-x)$

One of the most popular kernel function is an Epanechnikov kernel 0:

$$K(X) = 0.75(1 - x^2)$$

which is one of the representatives of the kernels class:

$$K(x) = \begin{cases} \frac{1}{v_d}(1 - x^2)^d \text{ for } x \in [-1, 1] \\ 0 \qquad\qquad \text{for } x \in (-\infty, -1) \cup (1, \infty) \end{cases}$$

B. *Smoothing Parameter*

It is very known in the literature that the selection of the smoothing parameter value has much bigger influence on the final regression results that the selection of a kernel function.

There exist a lot of methods of a smoothing parameter calculation. Method based on the approximation of the minimal regression error leads to two following result [12]:

$$h_0 = \left[\frac{R(K)}{\sigma_K^4 R(f'')}\right]^{0.2} n^{-0.2}$$

where $R(\blacksquare)$ and $\sigma_{\blacksquare}^t$ are the following function statistics:

$$R(K) = \int_{-\infty}^{\infty} K^2(x)dx$$

$$\sigma_K^t = \int_{-\infty}^{\infty} x^t K(x)dx$$

It causes some problems, because in the denominator of the fraction we need to know the form of the second derivative of the estimated (not known yet) function. Several

simplification (details can be found in [12]) lead to two formulas:

$$h_0 = 1.06\tilde{\sigma}n^{-0.2} \quad h_0 = 1.06\min(\tilde{\sigma}, 0.75\widehat{IQ})\,n^{-0.2}$$

where $\tilde{\sigma}$ is an estimator of the standard deviation of data sample x and $\widehat{IQ}$ is its interquartile range.

Amongst other method of smoothing parameter calculation maximal smoothing principle [15], cross validation methods [9]or nested methods [6] should be mentioned.

A lot of mentioned methods use the derivatives of kernel functions or kernel estimators of derivatives of estimated function. This leads to the problem of assuring kernel functions to be many times differentiable.

III. Spline Kernels

The main idea of introducing spline kernels is to make it possible to use the advantages of Epanechnikov (and other polynomial kernels) with the property of their differentiability. At each of a kernel domain an original kernel equation is replaced with another polynomial, which provides the final kernel continuity and differentiability.

Figure 1 presents the graphical illustration of this idea.

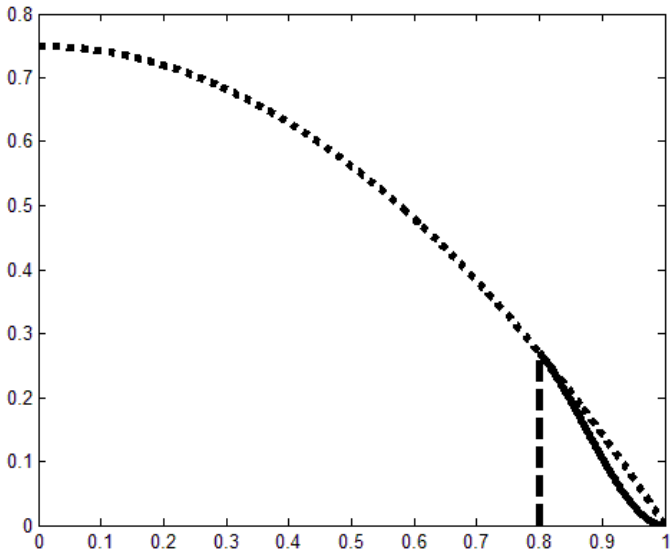

Figure 1. The idea of spline kernel with the cut at α=0.8: original Epanechnikov kernel (dotted), level of the cut (dashed) and the polynomial introduction (solid).

In this part of the paper two steps of defining a spline kernel are presented. First step (construction) consists of choosing the degree of the polynomial ant the point of the kernel splining (cut) while the second helps to assure one of the conditions of kernel function to be fulfilled.

A. *Construction*

As it was mentioned, one of the popular kernels family is as follows:

$$K(x) = \begin{cases} \frac{1}{v_d}(1 - x^2)^d \text{ for } x \in [-1,1] \\ 0 \qquad\qquad \text{for } x \in (-\infty,-1) \cup (1,\infty) \end{cases} \tag{1}$$

To assure the continuity and differentiability (demanded number of times) the equality of polynomial and kernel derivatives will be required. In the case of this family of kernels its derivative value can be expressed with the kernel as follows:

$$K'(x,d) = -\frac{1}{v_d}2dx(1 - x^2)^{d-1}$$
$$K'(x,d) = -\frac{2dx}{1-x^2 v_d}\left[\frac{1}{v_d}(1 - x^2)^d\right] \tag{2}$$
$$K'(x,d) = D(x,d)K(x,d)$$

With the assumption of the 3rd degree polynomial $g(x) = ax^3 + bx^2 + cx + d$ the following conditions must be satisfied:

$$\begin{cases} g(\alpha) = K(\alpha) & g'(\alpha) = K'(\alpha) \\ g(1) = 0 & g'(1) = 0 \end{cases} \tag{3}$$

which leads to the system of equations:

$$\begin{cases} \alpha^3 a + \alpha^2 b + \alpha c + d = K(\alpha) \\ 3\alpha^2 a + 2\alpha b + c + 0 = K'(\alpha) \\ a + b + c + d = 0 \\ 3a + 2b + c + 0 = 0 \end{cases}$$

The solution of this system of equations for the cut level $\alpha = 0.95$ (solved with the Matlab® software) is:

$$\begin{cases} a = 600.00 & c = 1681.50 \\ b = -1740.70 & d = -540.75 \end{cases}$$

At this point we have a definition of a spline kernel of this form:

$$K_M(X) = \begin{cases} 0 & |x| > 1 \\ K(x) & 0 \le x < \alpha \\ g(|x|) & \alpha \le |x| \le 1 \end{cases}$$

In the presented case its form can be presented as:

$$K_M(X) = \begin{cases} 0 & |x| > 1 \\ K(x) & 0 \le x < 0.95 \\ 600|x|^3 - 1745.75|x|^2 + & \\ +1681.5|x| - & \\ -540.75 & 0.95 \le |x| \le 1 \end{cases}$$

B. *Scaling*

Due to the modification of the kernel equation on the ends of its domain, it is no longer integrable to one. Then it requires a scaling parameter i, that is dependent of the α. The scaled kernel equation will be as follows:

$$K_{Mi}(X) = \begin{cases} 0 & |x| > 1 \\ i(\alpha)K(x) & 0 \le x < \alpha \\ i(\alpha)g(|x|) & \alpha \le |x| \le 1 \end{cases}$$

This scaling factor can be calculated from the following equation:

$$i(x) = \frac{0.5}{\int_0^{\alpha} K(x)dx + \int_{\alpha}^{1} g(x)dx}$$

For the once-differentiable kernel with the α=0.95 the scaling factor is:

$$i(\alpha_{0.95}) \approx \frac{0.5}{0.49968} = 1.000640$$

what leads to the final form of the once-differentiable spline kernel:

$$K_{Mi}(X) = \begin{cases} 0 & |x| > 1 \\ 1.0064\,K(x) & 0 \le x < 0.95 \\ 1.0064[600|x|^3 - & \\ -1745.75|x|^2 + 1681.5|x| - & \\ -540.75] & 0.95 \le |x| \le 1 \end{cases}$$

IV. Statistical Properties of Spline Kernels

Theorem 1: A spline kernel, generated on the basis of the kernel K, is an asymptotically statistically equivalent to the kernel K with $\alpha \to 1$.

As the statistical equivalence of two kernel functions, two conditions are considered, due to their influence on methods of smoothing parameter estimation: $R(K)$ and σ_K^2, which definitions are presented in the section II.B.

The proof of this theorem will be divided into two proofs – each one for a mentioned parameter. These proofs are based on the following lemmas ($g(x)$ is a polynomial):

Lemma 1. $\lim_{h\to 0} \int_t^{t+h} g(x)dx = 0 \quad t \in R$

Lemma 2. $\lim_{h\to 0} \int_t^{t+h} x^2 g(x)dx = 0 \quad t \in R$

Lemma 3. $\lim_{h\to 0} \int_t^{t+h} g(x)^2 dx = 0 \quad t \in R$

Lemma 4. $\lim_{\alpha\to 1} i(\alpha) = 1$

In proofs, numbers preceded with # marks the used lemma.

Theorem 1.1 – Thesis:

$$\lim_{\alpha\to 1} R(K_{Mi}, \alpha) = R(K)$$

Proof:

$$R(K_{Mi}, \alpha) = 2\int_0^1 K_{Mi}^2(x)dx = 2\int_0^{\alpha} i(\alpha)^2 K^2(x)dx + 2\int_{\alpha}^{1} i(\alpha)^2 g^2(x)dx$$

$$\lim_{\alpha\to 1} R(K_{Mi}, \alpha) = \lim_{\alpha\to 1} 2\underbrace{i(\alpha)^2}_{=1(\#4)}\left(\int_0^{\alpha} K^2(x)dx + \underbrace{\int_{\alpha}^{1} g^2(x)dx}_{=0\,(\#3)}\right)$$
$$= \lim_{\alpha\to 1} 2\int_0^{\alpha} K^2(x)dx = \int_{-1}^{1} K^2(x)dx$$
$$= R(K)$$

Theorem 1.2 – Thesis

$$\lim_{\alpha\to 1} \sigma^2(K_{Mi}, \alpha) = \sigma_K^2$$

Proof:

$$\sigma^2(K_{Mi}, \alpha) = 2\int_0^1 x^2 K_{Mi}(x)dx =$$

$$= 2\int_0^{\alpha} x^2 i(\alpha)^2 K(x)dx + 2\int_{\alpha}^{1} x^2 i(\alpha)^2 g(x)dx$$

$$\lim_{\alpha\to 1} \sigma^2(K_{Mi}, \alpha) =$$

$$= \lim_{\alpha\to 1}\left(2\int_0^{\alpha} x^2 i(\alpha)^2 K(x)dx + 2\int_{\alpha}^{1} x^2 i(\alpha)^2 g(x)dx\right) =$$

$$= \lim_{\alpha\to 1} 2\underbrace{i(\alpha)^2}_{=1(\#4)}\left(\int_0^{\alpha} x^2 K(x)dx + \underbrace{\int_{\alpha}^{1} x^2 g(x)dx}_{=0\,(\#2)}\right) =$$

$$= \lim_{\alpha\to 1} 2\int_0^{\alpha} x^2 K(x)dx = \sigma_K^2$$

V. Discussion and Conclusions

In this paper the new family of kernel functions was presented. The developed kernels are dedicated for methods of kernel regression which requires higher orders of kernel function derivatives.

Acknowledgement

This work was supported by the European Union from the European Social Fund (grant agreement number: UDA-POKL.04.01.01-106/09).

References

[1] W. S. Cleveland, and S. J. Devlin, "Locally weighted regression: An approach to regression analysis by local fitting", J. of the Am. Stat. Ass., vol. 83(403), pp. 596–610, 1988.

[2] C. de Boor, "A practical guide to splines", Springer, 2001.

[3] V. A. Epanechnikov, "Nonparametric estimation of a multivariate probability density", Theory of Probability and Its Applications, vol. 14, pp. 153 – 158, 1969.

[4] J. Fan, and I. Gijbels, "Variable bandwidth and local linear regression smoothers", Ann. Stat., vol. 20, pp .2008–2036, 1992.

[5] J. H. Friedman, "Multivariate adaptive regression splines", Ann. of Stat., vol. 19(1), pp. 1–141, 1991.

[6] T. Gasser, and A. Kneip, and K. Kohler, "A flexible and fast method for automatic smoothing", J. Am. Stat. Assoc., vol. 415, pp. 643–652, 1991

[7] T. Gasser, and H.G. Muller, "Estimating regression function and their derivatives by the kernel method", Scand. J. Stat., vol. 11, pp.171–185, 1984.

[8] T. Hastie, and R. Tibshirani, "Generalized additive models", Stat. Science, vol. 1(3), pp. 297 –318, 1986.

[9] J. S. Marron, "A comparison of cross-validation techniques in density estimation", Ann. of Stat., vol. 15(1), pp. 152–162, 1987.

[10] M. Michalak, and K. Nurzyńska, "Advanced Oblique Rule Generating Based on PCA", Lect. Notes in Comp. Sci., 2014 (in press)

[11] N. Nadaraya, "On Estimating Regression", Theory of Probab. and Its App., vol. 9(1), pp. 141–142, 1964.

[12] B.W. Silverman, "Density estimation for statistics and data analysis", Chapman & Hall, 1986.

[13] A. J. Smola, and B. Schoelkopf, "A tutorial on support vector regression", Statistics and Computing, vol. 14(3), pp.199–222, 2004.

[14] J. S. Taylor, and N. Cristianini, "Kernel Methods for Pattern Analysis", Cambridge University Press, 2004.

[15] G.R. Terrell, "The maximal smoothing principle in density estimation:, J. Am. Stat. Ass., vol. 410, pp. 470–477, 1990.

[16] V. N. Vapnik, "Statistical Learning Theory", Wiley, 1988.

[17] M.P. Wand, and M.C. Jones "Kernel smoothing", Chapman & Hall, 1995.

[18] G. Watson, "Smooth Regression Analysis". Sankhya - The Indian J. of Stat., vol 26(4), pp. 359–372, 1964.

About Author:

M. Michalak (1981) received his M.Sc. Eng. and Ph.D. degree in the field of computer science from the Silesian University of Technology (SUT), Poland in 2005 and 2009, respectively. Since 2009 he is a postdoctoral fellow in the Faculty of Automatic Control, Electronics and Computer Science at SUT and in the years 2009 – 2012 he was an engineer (later the Assistant Professor) at Central Mining Institute, Poland. He was also employed in the coal mining industry as the specialist of the data analysis. He is an author of over 50 publications from the fields of data mining, machines diagnosis, binary biclustering, rough sets theory, time series analysis, multi-spectral images analysis, oblique rule induction and many others.