ANNALS OF THE ORADEA UNIVERSITY
Fascicle of Management and Technological Engineering
ISSUE #1, MAY 2013, http://www.imtuoradea.ro/auo.fmte/

APPLICATION OF NON-PARAMETRIC REGRESSION IN ENGINEERING OPTIMIZATION

Imre KOCSIS[1], Tamas MANKOVITS[2]
[1]University of Debrecen Faculty of Engineering, kocsisi@eng.unideb.hu
[2] University of Debrecen Faculty of Engineering, tamas.mankovits@eng.unideb.hu

***Abstract*—**Non-parametric regression models (e.g. kernel smoothing techniques, support vector regression model) have been widely used in statistics and recently in econometrics and engineering as well. Regression functions can be effective tools in the solution of engineering optimization problems. In the investigation of complex technical systems the class of the function describing the connection between the input and output data is generally unknown, moreover the “classical” least square fitting method is not flexible enough to provide an effective regression function in the case of higher dimensional optimization problems. In this paper regression functions obtained in different models are compared graphically in one dimensional case. Calculations were made using Maple, R, and MS Excel software.

***Keywords*—**kernel function, regression

I. Introduction

THE smoothing techniques and the non-parametric approach have a long tradition in empirical analysis. The intensive interest in smoothing over the previous decades had two reasons: “statisticians realized that pure parametric thinking in curve estimations often does not meet the need for flexibility in data analysis and the development of hardware created the demand for theory of now computable nonparametric estimates.” [1]

Kernel smoothing is an effective regression technique for engineering optimization as well. It has the advantage that the function defining the non-linear relation between the input and output data does not need to be given in explicit form. As an example the shape optimization problem of rubber bumpers can be mentioned that was investigated by the second author in his doctoral thesis [7].

To present the flexibility of the kernel smoothing and the support vector regression we give the regression function for certain input data (one dimensional case).

II. Kernel functions

Let us consider a data set (frequently called training date set)

$$\left\{(\bar{\mathbf{t}}_1,\mathbf{d}_1),\dots,(\bar{\mathbf{t}}_P,\mathbf{d}_P)\right\}\subset \mathbf{R}^{N+1}, \tag{1}$$

where $\bar{t}_i$ are (possibly one dimensional) input vectors, d_i are associated targets (R is the set of reals). The goal is to fit a function $f: R^N \to R$ which approximates the relation between the data sets. Any regression algorithm has a loss function $L\left(f\left(\bar{t}\right),d\right)$ which describes how the estimated function deviated from the true one. In this note we refer to the following two types of loss function: the quadratic loss function (least square fitting)

$$\mathbf{L}\left(\mathbf{f}\left(\bar{\mathbf{t}}_i\right),\mathbf{d}_i\right)=\left(\mathbf{f}\left(\bar{\mathbf{t}}_i\right)-\mathbf{d}_i\right)^2 \tag{2}$$

and the ε-insensitive (Vapnik’s) loss function

$$\mathbf{L}\left(\mathbf{f}\left(\bar{\mathbf{t}}_i\right),\mathbf{d}_i\right)=\begin{cases}\mathbf{0} & \textbf{if } \left|\mathbf{f}\left(\bar{\mathbf{t}}_i\right)-\mathbf{d}_i\right|\le\boldsymbol{\varepsilon}\\ \left|\mathbf{f}\left(\bar{\mathbf{t}}_i\right)-\mathbf{d}_i\right|-\boldsymbol{\varepsilon} & \textbf{otherwise}\end{cases}, \tag{3}$$

where ε is a fixed positive parameter.

It is well-know that in the linear model (with the quadratic loss function)

$$\sum_{\mathbf{i}=1}^{\mathbf{P}}\left(\bar{\mathbf{w}}\cdot\bar{\mathbf{t}}_i+\mathbf{b}-\mathbf{d}_i\right)^2\xrightarrow[\bar{\mathbf{w}}]{}\mathbf{min} \tag{4}$$

the regression function can be expressed as a linear combination of so-called kernel functions

$$\mathbf{k}_i\left(\bar{\mathbf{x}}\right)=\bar{\mathbf{x}}\cdot\bar{\mathbf{t}}_i \tag{5}$$

in the form

$$\mathbf{f}\left(\bar{\mathbf{x}}\right)=\bar{\mathbf{w}}^{\mathbf{opt}}\cdot\bar{\mathbf{x}}+\mathbf{b}=\sum_{\mathbf{i}=1}^{\mathbf{P}}\alpha_i\mathbf{k}_i\left(\bar{\mathbf{x}}\right)+\mathbf{b}, \tag{6}$$

where $\cdot$ denotes the inner product [5].

In non-linear models the situation is similar: applying a suitable transformation $\bar{\phi}: R^N \to R^M$ the problem will be linear in R^M and the regression function is a linear combination of kernel functions

$$\mathbf{k}_i\left(\bar{\mathbf{x}}\right)=\bar{\phi}\left(\bar{\mathbf{x}}\right)\cdot\bar{\phi}\left(\bar{\mathbf{t}}_i\right) \tag{7}$$

[3], [4] in the form

$$\mathbf{f}\left(\bar{\mathbf{x}}\right)=\bar{\mathbf{w}}^{\mathbf{opt}}\cdot\bar{\phi}\left(\bar{\mathbf{x}}\right)+\mathbf{b}=\sum_{\mathbf{i}=1}^{\mathbf{P}}\alpha_i\mathbf{k}_i\left(\bar{\mathbf{x}}\right)+\mathbf{b}. \tag{8}$$

Since function $\bar{\phi}$ and, consequently, functions k_i are generally unknown in practice, one of the recommended kernel functions is chosen in the model to build f. The most used kernel in technical literature is the radial base Gaussian kernel function

$$\mathbf{k_i(\bar{x}) = e^{-\gamma\cdot\|\bar{x}-\bar{t}_i\|^2}}, \tag{9}$$

where γ is a parameter determining the shape of the kernel function. The role of γ is presented graphically in section III.

III. Application of Gaussian kernel function

As an input data set consider the set of the following points *{(t*$_i$*,d*$_i$*)|i=1,..,P}={(1,16),(2,19),(3,9),(4,16),(5,25),(6,10),(7,5),(8,4),(9,24),(10,20)}* (Fig.5.) and solve the quadratic optimization problem

$$\mathbf{f(x) = \sum_{i=1}^{P}\alpha_i k_i(x) + b = \sum_{i=1}^{P}\alpha_i \cdot e^{-\gamma|x-t_i|^2} + b}$$

$$\mathbf{\sum_{i=1}^{P}(f(t_i)-d_i)^2 \xrightarrow[\bar{\alpha},b]{} min} \tag{10}$$

for $\bar{\alpha} = (\alpha_1,\ldots,\alpha_P) \in R^P$ and $b \in R$, where γ is a fixed positive number.

The following figures show graphically the regression functions obtained for certain values of γ. (Calculations and plotting were carried out using Maple).

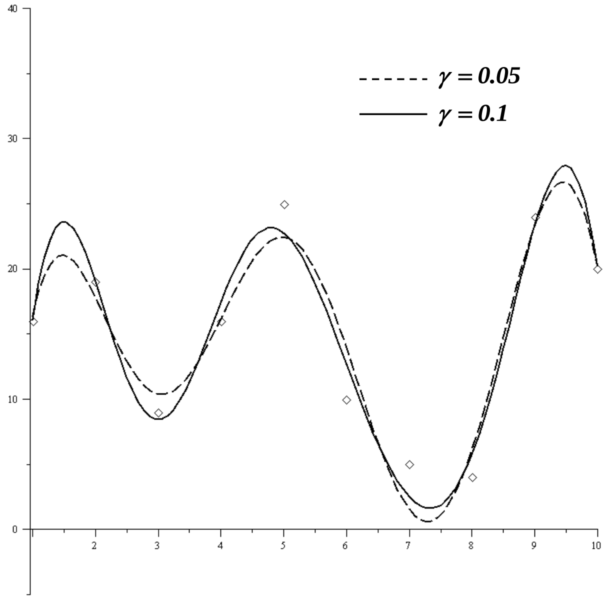

Fig. 1. The regression functions (solutions of the optimization problem (10)) when *γ=0.05* and *γ=0.1*

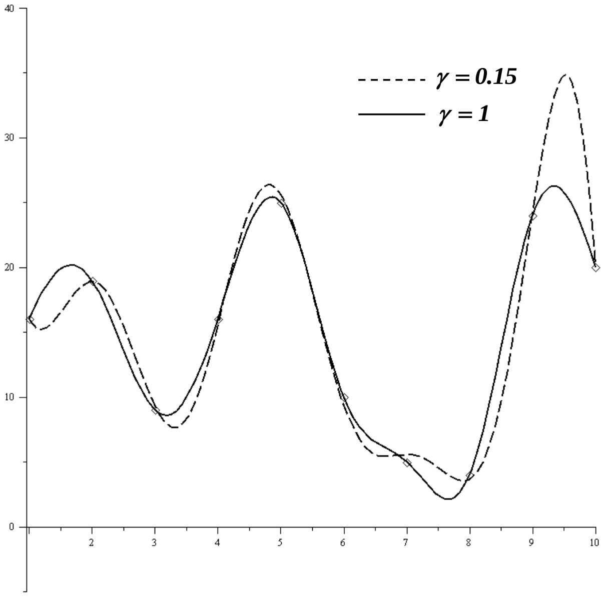

Fig. 2. The regression functions (solutions of the optimization problem (10)) when *γ=0.15* and *γ=1*

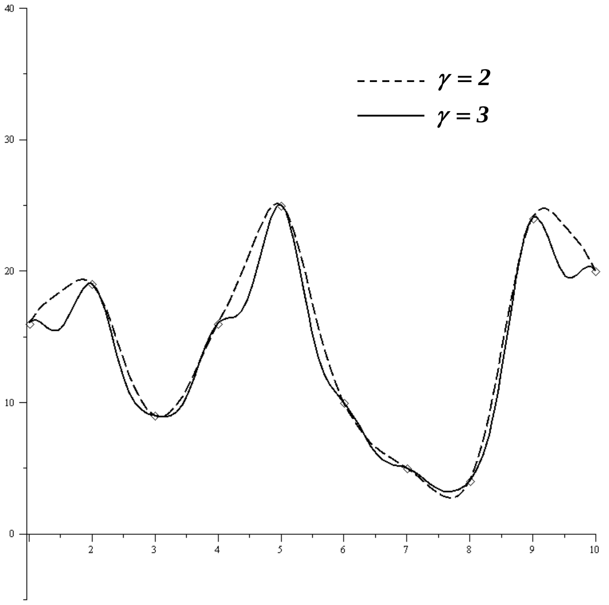

Fig. 3. The regression functions (solutions of the optimization problem (10)) when *γ=2* and *γ=3*

It can be seen that the smaller values of γ result flatter regression functions (see e.g. Fig.1.) but the deviation of these "flat" functions at the learning points can be relatively high. Large values of γ lead to regression functions varying too fast and the shape of the regression functions do not seem to be reasonable (see e.g. Fig.4.). Analyzing the pictures it can be seen that the values $\gamma = 1..2$ mean an acceptable compromise between the flatness of the function and the accuracy at the learning points (Fig.2., Fig.3.).

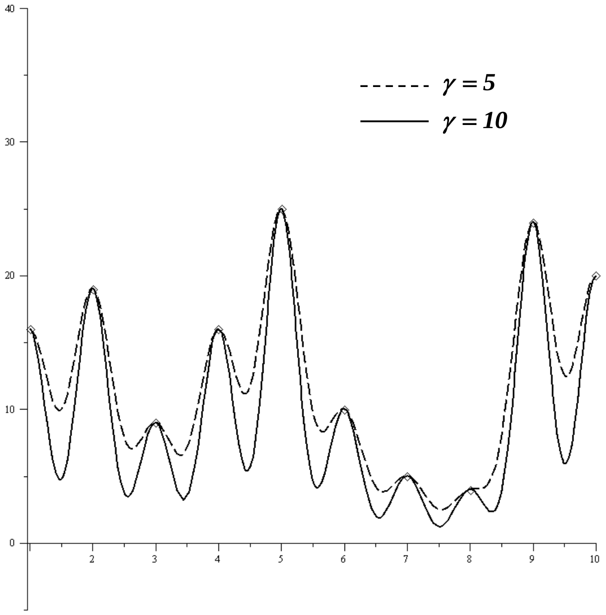

Fig. 4. The regression functions (solutions of the optimization problem (10)) when γ=5 and γ=6

IV. SUPPORT VECTOR REGRESSION MODEL

Application of the ε -insensitive loss function (3) leads to a quadratic convex optimization problem and the regression function can be expressed by kernel functions.

In the linear model the regression function is sought in the form

$$\mathbf{f(\bar{x}) = \bar{w}\cdot\bar{x} + b, \quad \bar{w} \in R^N, b \in R}. \tag{11}$$

The flatness of f in this case means that one seeks a small $\bar{w}$, that is, the norm $\|\bar{w}\|$ is to be minimized. The related convex optimization problem is the following:

$$\begin{array}{ll} \textbf{minimize} & \frac{1}{2}\|\bar{\mathbf{w}}\|^2 \\ \textbf{subject to} & \begin{cases} \mathbf{d_i - (\bar{w}\cdot\bar{t}_i + b) \le \varepsilon} \\ \mathbf{(\bar{w}\cdot\bar{t}_i + b) - d_i \le \varepsilon} \end{cases}, \quad \mathbf{i = 1,\ldots,P.} \end{array} \tag{12}$$

Inequalities in (12) mean that the regression function is required to approximate the pairs $(\bar{t}_i, d_i)$ with ε precision. Sometimes the convex optimization problem is infeasible with the given constraints and some “errors” must be allowed. For this purpose so-called slack variables was introduced by Vapnik in [2]. The modified convex optimization problem with slack variables ξ_i and ξ_i' is the following

$$\begin{array}{ll} \textbf{minimize} & \frac{1}{2}\|\bar{\mathbf{w}}\|^2 + \mathbf{C}\sum_{i=1}^{P}(\xi_i + \xi_i') \\ \textbf{subject to} & \begin{cases} \mathbf{d_i - (\bar{w}\cdot\bar{t}_i + b) \le \varepsilon + \xi_i} \\ \mathbf{(\bar{w}\cdot\bar{t}_i + b) - d_i \le \varepsilon + \xi_i'}, \quad \mathbf{i = 1,\ldots,P.} \\ \mathbf{\xi_i \ge 0, \xi_i' \ge 0} \end{cases} \end{array} \tag{13}$$

The positive constant C determines the trade-off between the flatness of f and the amount up to which deviations larger than ε are tolerated.

According to the standard dualization method [6] we introduce the Lagrange function

$$\mathbf{L(\bar{w}, b, \bar{\lambda}_i, \bar{\lambda}_i', \bar{\alpha}_i, \bar{\alpha}_i') = \frac{1}{2}\|\bar{w}\|^2 + C\sum_{i=1}^{P}(\xi_i + \xi_i')}$$
$$\mathbf{-\sum_{i=1}^{P}(\lambda_i\xi_i + \lambda_i'\xi_i') - \sum_{i=1}^{P}\alpha_i(\varepsilon + \xi_i + (\bar{w}\cdot\bar{x} + b) - d_i)} \tag{14}$$
$$\mathbf{-\sum_{i=1}^{P}\alpha_i'(\varepsilon + \xi_i' + d_i - (\bar{w}\cdot\bar{x} + b))}$$

where $\alpha_i, \alpha_i', \lambda_i, \lambda_i' \ge 0$. The dual optimization problem is the following:

$$\begin{array}{ll} \textbf{maximize} & \mathbf{-\frac{1}{2}\sum_{i,j=1}^{P}(\alpha_i - \alpha_i')(\alpha_j - \alpha_j')\bar{t}_i\cdot\bar{t}_j} \\ & \mathbf{-\varepsilon\sum_{i=1}^{P}(\alpha_i + \alpha_i') + \sum_{i=1}^{P}(\alpha_i - \alpha_i')\bar{t}_i} \\ \textbf{subject to} & \mathbf{\sum_{i=1}^{P}(\alpha_i - \alpha_i') = 0 \ \text{and} \ \alpha_i, \alpha_i' \in [0, C].} \end{array} \tag{15}$$

Solving the dual problem we get

$$\mathbf{f(\bar{x}) = \sum_{i=1}^{P}(\alpha_i - \alpha_i')\bar{x}\cdot\bar{t}_i + b}, \tag{16}$$

that is, the solution is a linear combination of kernel functions. This form of f says that the explicit form of $\bar{w}$ does not need to be computed. Furthermore it can be proved that for the learning points inside the ε-tube $\alpha_i - \alpha_i' = 0$, that is, f is determined by the learning points having nonvanishing coefficients. These pairs are called support vectors.

As it mentioned before, in non-linear model the application of a suitable transformation function $\bar{\phi}: R^N \to R^M$ leads to a linear problem. It can be proved that in this case regression function is a linear combination of kernel functions $k_i(\bar{x}) = \bar{\phi}(\bar{x})\cdot\bar{\phi}(\bar{t}_i)$.

The convex optimization problem is the following

$$\begin{array}{ll} \textbf{minimize} & \frac{1}{2}\|\bar{\mathbf{w}}\|^2 + \mathbf{C}\sum_{i=1}^{P}(\xi_i + \xi_i') \\ \textbf{subject to} & \begin{cases} \mathbf{d_i - (\bar{w}\cdot\phi(\bar{t}_i) + b) \le \varepsilon + \xi_i} \\ \mathbf{(\bar{w}\cdot\phi(\bar{t}_i) + b) - d_i \le \varepsilon + \xi_i', i = 1,\ldots,P.} \\ \mathbf{\xi_i \ge 0, \xi_i' \ge 0} \end{cases} \end{array} \tag{17}$$

The dual problem is

$$\begin{array}{ll} \textbf{maximize} & \mathbf{-\frac{1}{2}\sum_{i,j=1}^{P}(\alpha_i - \alpha_i')(\alpha_j - \alpha_j')\bar{\phi}(\bar{t}_i)\cdot\bar{\phi}(\bar{t}_j)} \\ & \mathbf{-\varepsilon\sum_{i=1}^{P}(\alpha_i + \alpha_i') + \sum_{i=1}^{P}(\alpha_i - \alpha_i')\bar{t}_i} \\ \textbf{subject to} & \mathbf{\sum_{i=1}^{P}(\alpha_i - \alpha_i') = 0 \ \text{and} \ \alpha_i, \alpha_i' \in [0, C].} \end{array} \tag{18}$$

Solving the dual problem we get

$$\mathbf{f(\bar{x}) = \sum_{i=1}^{P}(\alpha_i - \alpha_i')\bar{\phi}(\bar{x})\cdot\bar{\phi}(\bar{t}_i) + b}, \tag{19}$$

that is, the solution is a linear combination of kernel functions.

Using the SVR package of the R software we present the role of parameters ε, γ and C in the SVR model. We consider the same input data set as in section II: *{(t_i,d_i)|i=1,..,P}={(1,16),(2,19),(3,9),(4,16),(5,25),(6,10),(7,5),(8,4),(9,24),(10,20)}* (Fig.5.) and choose the Gaussian kernel function (9). (Plotting was carried out using MS Excel.)

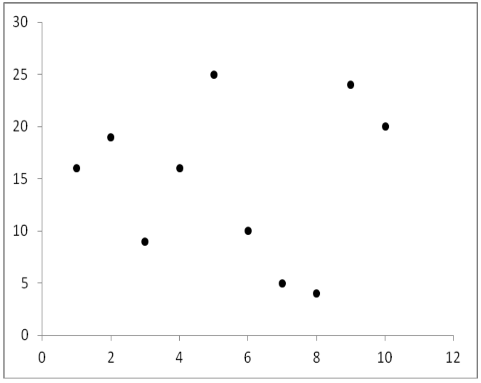

Fig. 5. The set of input data

Fig.6. shows the role of the "penalty" parameter C. Application of a higher value of C results a regression function with small deviation even in outlying learning points, while low value of C gives a flat function with large deviation at certain points.

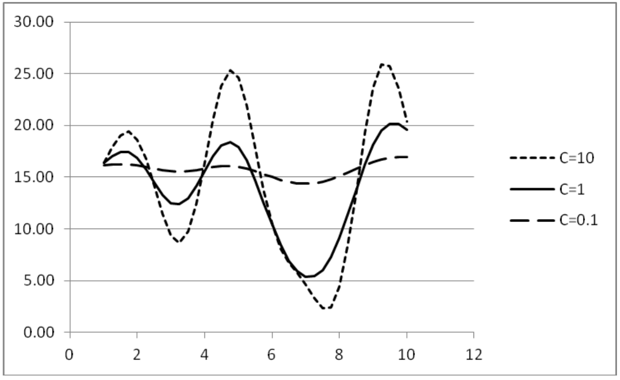

Fig. 6. The role of the "penalty" parameter *C* (*C*=0.1,1,10, γ=5, ε=0.05)

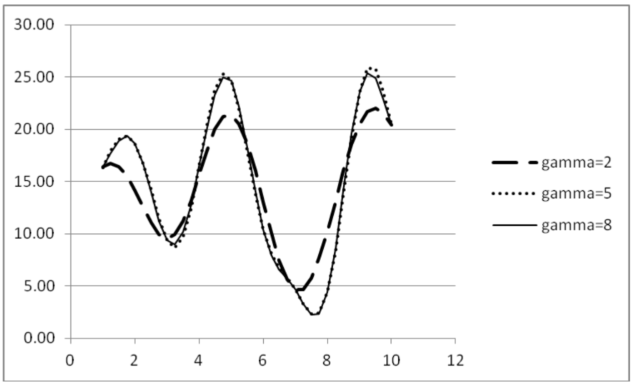

Fig. 7. The role of the parameter γ (γ=2,5,8, *C*=10, ε=0.05)

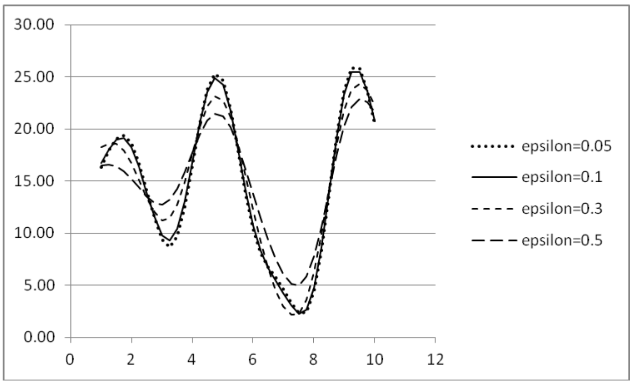

Fig. 8. The role of the parameter ε (ε=0.05,0.1,0.3,0.5, γ=5, *C*=10)

Fig.7. shows the role of the parameter γ (it was described also in section II). Fig.8. shows the role of the parameter ε. The learning points are possibly (depending on the value of the slack variables) in an ε-tube or near the ε-tube.

Examples presented in this note show the flexibility of the regression methods using kernel functions. In engineering applications this flexibility can have an important role when the model has to be adjusted to the special characteristics of the input data set or to the requirements related to the regression function.

ACKNOWLEDGMENT

The described work was carried out as part of the TÁMOP-4.2.2/B-10/1-2010-0008 and TÁMOP-4.2.2.A-11/1/KONV-2012-0041 projects in the framework of the New Hungarian Development Plan. The realization of this project is supported by the European Union, co-financed by the European Social Fund.

REFERENCES

[1] W. Hardle, "Applied non-parametric regression", *Humbolt-Universität*, Berlin, 1994
[2] C. Cortes and V. Vapnik, "Support vector networks", *Machine Learning*, Vol. 20, pp. 273-297, 1995.
[3] B. Schölkopf, and A. J. Smola, "Learning with Kernels", *MIT Press*, 2000
[4] A. Farag and R. M. Mohamed, "Regression Using Support Vector Machines: Basic Foundations", *University of Louisville*, 2004
[5] S. Haykin, "Neural Networks and Learning Machines", *Prentice Hall*, 2009
[6] S. Boyd and L. Vandenberghe, "Convex Optimization", *Cambridge University Press*, 2009
[7] T. Mankovits, "Shape optimization of rubber parts", PhD Thesis, *University of Miskolc*, 2013