On Learning in a Time-Varying Environment by Using a Probabilistic Neural Network and the Recursive Least Squares Method

Maciej Jaworski and Marcin Gabryel

Department of Computer Engineering, Czestochowa University of Technology, Armii Krajowej 36, 42-200 Czestochowa, Poland
{maciej.jaworski,marcin.gabryel}@kik.pcz.pl

Abstract. This paper presents the recursive least squares method, combined with the general regression neural networks, applied to solve the problem of learning in time-varying environment. The general regression neural network is based on the orthogonal-type kernel functions. The appropriate algorithm is presented in a recursive form. Sufficient simulations confirm empirically the convergence of the algorithm.

1 Introduction

The idea of probabilistic neural networks and general regression neural networks was first proposed by Specht in [35] and [36], respectively. Such networks are nonparametric tools, designed for estimating probability density and regression functions. In literature, their usability in solving stationary (see e.g. [4], [5], [7], [12]-[16], [22]-[24] and [27]-[30]) and nonstationary problems (see e.g. [8], [17]-[21], [25] and [26]) has been widely studied. It should be emphasized that in both cases the noise was assumed to be stationary. An excellent overwiew of the methods mentioned above can be found in [6] and [9].

Let us consider a system, which processes p-dimensional data elements $X_i \in A \subset \mathbb{R}^p$, $i = 1, \ldots$, with some unknown function $\phi : A \to \mathbb{R}$ ($E[\phi(X_i)] < \infty$). The probability density function of the random variables X_i, described by $f(x)$, is unknown as well. Let us assume that the output $\phi(X_i)$ of the system is accompanied with a noise, consisting of two components

- deterministic part ac_i, where a is some unknown constant, and c_i is an element of known sequence, satisfying $\lim_{i\to\infty} |c_i| = \infty$,
- probabilistic part Z_i, which is a random variable satisfying the following condition

$$E[Z_i] = 0, \quad E[Z_i^2] = d_i < \infty. \tag{1}$$

Therefore, the output random variable Y_i, received from the system, is given by the equation

$$Y_i = \phi(X_i) + ac_i + Z_i. \tag{2}$$

L. Rutkowski et al. (Eds.): ICAISC 2012, Part I, LNCS 7267, pp. 99–110, 2012.

The aim of the generalized nonlinear regression is to estimate simultaneously the regression function $\phi(x)$ and the constant a, given n pairs of random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$ and the sequence c_i.

2 Estimation of the Parameter *a*

For the estimation of the parameter a, the recursive least square method [1] can be applied

$$\hat{a}_n = \hat{a}_{n-1} + \frac{c_n}{\sum_{i=1}^{n} c_i^2} \left(Y_n - \hat{a}_{n-1} c_n \right). \tag{3}$$

This method can be further generalized into the form

$$\hat{a}_n^{(\omega)} = \hat{a}_{n-1}^{(\omega)} + \frac{c_n^{\omega-1}}{\sum_{i=1}^{n} c_i^{\omega}} \left(Y_n - \hat{a}_{n-1}^{(\omega)} c_n \right), \tag{4}$$

where ω is a real nonnegative number.

The assumptions of the following theorem ensure the convergence of the estimator $\hat{a}_n^{(\omega)}$ to the actual value of parameter a

Theorem 1. *If conditions (1) holds and additionally the following conditions are satisfied*

$$E[\phi^2(X_i)] = \int_A \phi^2(x) f(x) dx < \infty, \tag{5}$$

$$\lim_{n \to \infty} \left(\frac{\sum_{i=1}^{n} c_i^{\omega-1}}{\sum_{i=1}^{n} c_i^{\omega}} \right) = 0, \tag{6}$$

$$\lim_{n \to \infty} \left(\frac{\sum_{i=1}^{n} c_i^{2\omega-2} s_i}{\left(\sum_{i=1}^{n} c_i^{\omega} \right)^2} \right) = 0, \tag{7}$$

where s_i is defined as follows

$$s_i = \max\{Var[\phi(X_i)], d_i\}, \tag{8}$$

then

$$\hat{a}_n^{(\omega)} \xrightarrow{n \to \infty} a \ \textit{ in probability}. \tag{9}$$

Proof. The theorem can be proven using simple analysis of the bias and the variance of estimator (4), which leads to the following convergence

$$\lim_{n \to \infty} E\left[\left(\hat{a}_n^{(\omega)} - a \right)^2 \right] = 0. \tag{10}$$

Convergence (10) is the sufficient condition for convergence (9).

3 Estimation of the Regression Function $\phi(x)$

To find the regression function $\phi(x)$, the nonlinear regression procedures should be applied to the pairs of random variables (X_i, V_i), where

$$V_i = Y_i - ac_i. \tag{11}$$

Since the actual value of the parameter a is not known, the random variables V_i have to be estimated, using some estimator $\overline{a}_i$

$$\hat{V}_i(\overline{a}_i) = Y_i - \overline{a}_i c_i,\ i = 1, \ldots, n. \tag{12}$$

It is easily seen that if $\overline{a}_i = a$, then $\hat{V}_i(\overline{a}_i) = \hat{V}_i(a) \equiv V_i$. In this section, for further considerations of the convergence of the regression function estimator, it is assumed that $\overline{a}_i = a,\ i = 1, \ldots, n$.

The regression function $\phi(x)$ can be expressed in the following form

$$\phi(x) = \frac{\phi(x)f(x)}{f(x)} \stackrel{def.}{=} \frac{R(x)}{f(x)}, \tag{13}$$

at each point x, for which $f(x) \neq 0$. The nominator and the denominator of the above expression can be estimated separately. In this paper the nonparametric estimation based on kernel functions is proposed. Given $2n$ kernel functions $\tilde{K}_i, \tilde{K}'_i : A \times A \to \mathbb{R},\ i = 1, \ldots, n$, the estimators $\tilde{R}_n(x, a)$ and $\tilde{f}_n(x)$ can be expressed in the following form

$$\tilde{R}_n(x, a) = \frac{1}{n} \sum_{i=1}^{n} \hat{V}_i(a) \tilde{K}_i(x, X_i), \tag{14}$$

$$\tilde{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \tilde{K}'_i(x, X_i), \tag{15}$$

One way of constructing the kernel functions is the application of orthogonal series. Let $g_j : \mathbb{R} \to \mathbb{R},\ j = 1, \ldots$ denote the functions of the complete orthogonal system, satisfying the following condition

$$\forall_{j \in \mathbb{N}} \sup_{w \in \mathbb{R}} |g_j(w)| \leq G_j. \tag{16}$$

Then the kernel functions K, K', for one-dimensional case ($p = 1$), can be defined as follows

$$\tilde{K}_i(x, u) = \sum_{j=0}^{M(i)} g_j(x) g_j(u),\ \tilde{K}'_i(x, u) = \sum_{j=0}^{N(i)} g_j(x) g_j(u),\ i = 1, \ldots, n, \tag{17}$$

where $M(i)$ and $N(i)$ are sequences satisfying $\lim_{i \to \infty} M(i) = \infty$ and $\lim_{i \to \infty} N(i) = \infty$, respectively. The convergence of estimators (14) and (15) with kernel functions (16) can be slightly improved by the applicaion of so-called Cesaro means. Let us denote $S_j(x,u)$ as the following partial sums

$$S_j(x,u) = \sum_{k=0}^{j} q_k(x)g_k(u),\ j = 1,\ldots,n. \tag{18}$$

Then, the kernel functions $K, K' : A \times A \to \mathbb{R}$ can be proposed as the Cesaro means of these partial sums

$$K_i(x,u) = \frac{1}{M(i)+1}\sum_{j=0}^{M(i)} S_j(x,u) = \sum_{j=0}^{M(i)}\left(1 - \frac{j}{M(i)+1}\right) g_j(x)g_j(u),\ i = 1,\ldots,n, \tag{19}$$

$$K'_i(x,u) = \frac{1}{N(i)+1}\sum_{j=0}^{N(i)} S_j(x,u) = \sum_{j=0}^{N(i)}\left(1 - \frac{j}{N(i)+1}\right) g_j(x)g_j(u),\ i = 1,\ldots,n \tag{20}$$

Finally, the estimator for functions $R(x)$ and $f(x)$ can be proposed in the following forms

$$\overline{R}_n(x,a) = \frac{1}{n}\sum_{i=1}^{n} \hat{V}_i(a)K_i(x,X_i) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{M(i)} \hat{V}_i(a)\left(1 - \frac{j}{M(i)+1}\right) g_j(x)g_j(X_i), \tag{21}$$

$$\overline{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K'_i(x,X_i) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{N(i)}\left(1 - \frac{j}{N(i)+1}\right) g_j(x)g_j(X_i), \tag{22}$$

Then, in light of formula (13), the estimator of the regression function $\phi(x)$ is given by

$$\overline{\phi}_n(x,a) = \frac{\overline{R}_n(x,a)}{\overline{f}_n(x)}. \tag{23}$$

To ensure the convergence of estimator (23), assumptions of the following theorem have to be satisfied.

Theorem 2. *If conditions (1) and (5) hold and additionally the following conditions are satisfied*

$$\lim_{n\to\infty} N(n) = 0,\quad \lim_{n\to\infty}\left[\frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j=1}^{N(i)} G_j^2\right)^2\right] = 0, \tag{24}$$

$$\lim_{n\to\infty} M(n) = 0, \quad \lim_{n\to\infty} \left[\frac{1}{n^2} \sum_{i=1}^{n} \left(\sum_{j=1}^{M(i)} G_j^2 \right)^2 s_i \right] = 0, \tag{25}$$

where s_i is defined as in (8), then

$$\overline{\phi}_n(x,a) \overset{n\to\infty}{\longrightarrow} \phi(x) \text{ in probability.} \tag{26}$$

Proof. The proof of the theorem can be found in [14].

4 Probabilistic Neural Network

In the real world applications, the value of the parameter a is not known, therefore the random variables V_i have to be estimated, using formula (12). In particular, the estimators $\overline{a}_i$ can be the same for each variable V_i, e.g.

$$\forall_{i\in\{1,\ldots,n\}}\, \overline{a}_i = \hat{a}_n^{(\omega)}. \tag{27}$$

Then, the estimator of the regression function $\phi(x)$ can be proposed as $\overline{\phi}_n(x, \hat{a}_n^{(\omega)})$ (replacing a by $\hat{a}_n^{(\omega)}$ in formulas (21) and (23)). The convergence of this estimator can be proven combining Theorems 1 and 2 and Theorem 4.3.8 in [40]. The estimator $\overline{\phi}_n(x, \hat{a}_n^{(\omega)})$ is calculated in a two-step process. First, given all output variables Y_i, $i = 1, \ldots, n$, the estimator $\hat{a}_n^{(\omega)}$ is computed. Then, after the first step is completed, the estimation of the regression function $\phi(x)$ can be provided.

The main disadvantage of the approach presented above is that the algorithm cannot be performed in a recursive way. To maintain this ability, we propose a slightly modificated form of estimator (21) (and in consequence (23)). Let us assume that for each $i = 1, \ldots, n$ an estimator $\overline{a}_n$ is different end equal to $\hat{a}_n^{(\omega)}$. Then, the estimator of function $R(x)$ can be proposed as follows

$$\hat{R}_n(x, \{\hat{a}_i^{(\omega)}\}_n) = \frac{1}{n} \sum_{i=1}^{n} \hat{V}_i(\hat{a}_i^{(\omega)}) K_i(x, X_i), \tag{28}$$

where $\{\hat{a}_i^{(\omega)}\}_n$ denotes the subset of estimators $\{\hat{a}_1^{(\omega)}, \ldots, \hat{a}_n^{(\omega)}\}$ and K_i is the kernel function given by (19). Estimator (28) can be easily written in a recursive way

$$\hat{R}_n(x, \{\hat{a}_i^{(\omega)}\}_n) = \frac{n-1}{n} \hat{R}_{n-1}(x, \{\hat{a}_i^{(\omega)}\}_{n-1}) + \frac{1}{n} \hat{V}_n(\hat{a}_n^{(\omega)}) K_n(x, X_n), \tag{29}$$

where estimators $\hat{a}_n^{(\omega)}$ are computed using recursive formula (4). Estimator (22) can be expressed in a recursive way without any additional modifications

$$\overline{f}_n(x) = \frac{n-1}{n} \overline{f}_{n-1}(x) + \frac{1}{n} K_n'(x, X_n). \tag{30}$$

Finally, the estimator of the regression function $\phi(x)$ is, analogously to (23), given by

$$\hat{\phi}_n(x, \{\hat{a}_i^{(\omega)}\}_n) = \frac{\hat{R}_n(x, \{\hat{a}_i^{(\omega)}\}_n)}{\overline{f}_n(x)}. \quad (31)$$

The algorithm presented above can be considered as a general regression neural network [36]. The appropriate scheme of this network is presented in Fig. 1.

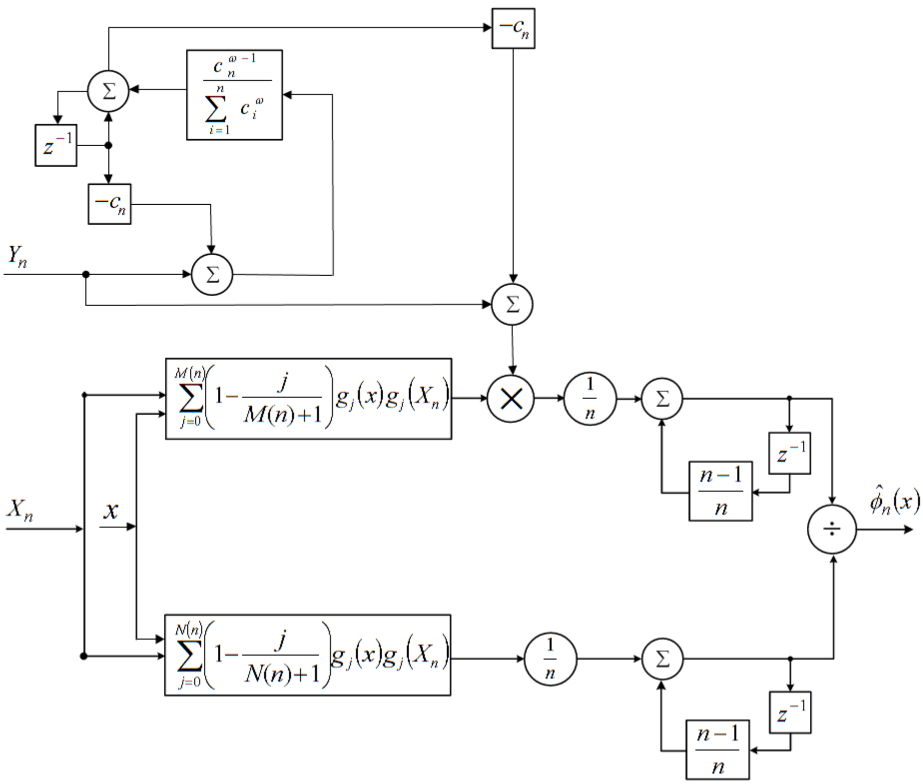

Fig. 1. The block digram of the probabilistic neural network, adopted to performing algorithms presented in sections 2 and 3

We do not present any theorem, which would ensure the convegence of estimator (31). Instead, in the next section the convergence is tested empirically, on a basis of several numerical simulations.

5 Simulations

In the following simulations a system described by equation (2) is considered, with the constant a equal to $2,5$ and the regression function $\phi(x)$ given by

$$\phi(x) = 10\frac{2x^3 - x}{\cosh(2x)}. \tag{32}$$

The random variables X_i are generated from the uniform probability distribution, from the interval $X_i \in [-5,5]$, $i = 1,\ldots,n$. The random variables Z_i come from the normal distribution $N(0,d_i)$, where d_i is given in the form

$$d_i = i^{\alpha},\ i = 1,\ldots,n,\ \ \alpha > 0. \tag{33}$$

The elements of the sequence c_i are taken in a similar form

$$c_i = i^{t},\ i = 1,\ldots,n,\ \ t > 0. \tag{34}$$

In the presented simulations, the parameter t is set to $t = 0,2$. It is easily seen that, in order to obey assumptions 6) and (7) of Theorem 1, the exponent α has to satisfy the following inequality

$$\alpha < 2t + 1. \tag{35}$$

In the estimators of functions $R(x)$ and $f(x)$, the Hermite orthogonal system is proposed

$$g_j(x) = \begin{cases} \dfrac{\exp\left(-x^2/2\right)}{\sqrt[4]{\pi}} & j = 0, \\ -\sqrt{2}x g_0(x) & j = 1, \\ -\sqrt{\dfrac{2}{j}}x g_{j-1}(x) - \sqrt{\dfrac{j-1}{j}}g_{j-2}(x) & j > 0. \end{cases} \tag{36}$$

The functions g_j can be bounded by (see [39])

$$\forall_{j\in\mathbb{N}} \sup_{x\in\mathbb{R}} |g_j(x)| \le G_j = Cj^{-1/12}. \tag{37}$$

Assuming that the sequences $M(n)$ and $N(n)$ are given in the following forms

$$M(n) = \lceil Dn^{Q} \rceil,\ D > 0, Q > 0,\ \ N(n) = \lceil D'n^{Q'} \rceil,\ D' > 0, Q' > 0, \tag{38}$$

the assumptions (24) and (25) of Theorem 2 are satisfied if the parameters Q and Q' satisfy the following conditions

$$Q' < \frac{3}{5},\ \ Q < \frac{3}{5}\left(1 - \alpha\right). \tag{39}$$

In all of the simulations, parameters Q and Q' are kept the same, i.e. $Q = Q'$. Parameters D and D' are set to $D = D' = 1,5$.

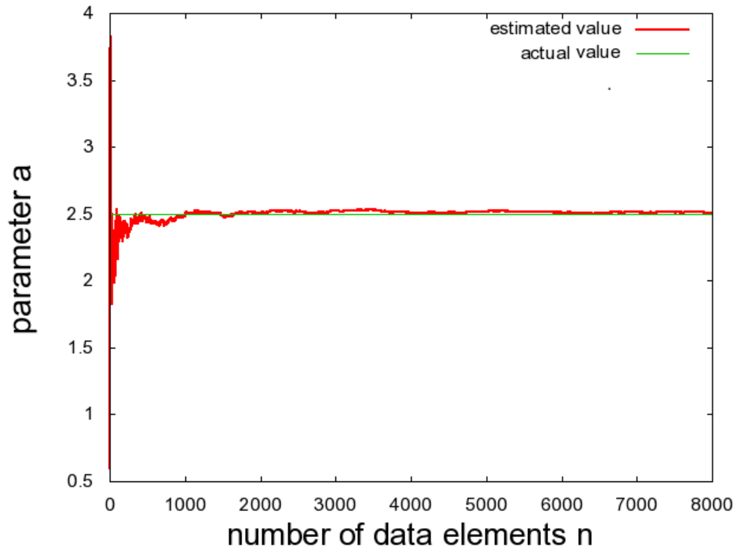

Fig. 2. Convergence of the estimator $\hat{a}_n^{(8)}$ to the actual value of parameter a, for $\alpha = 0,2$

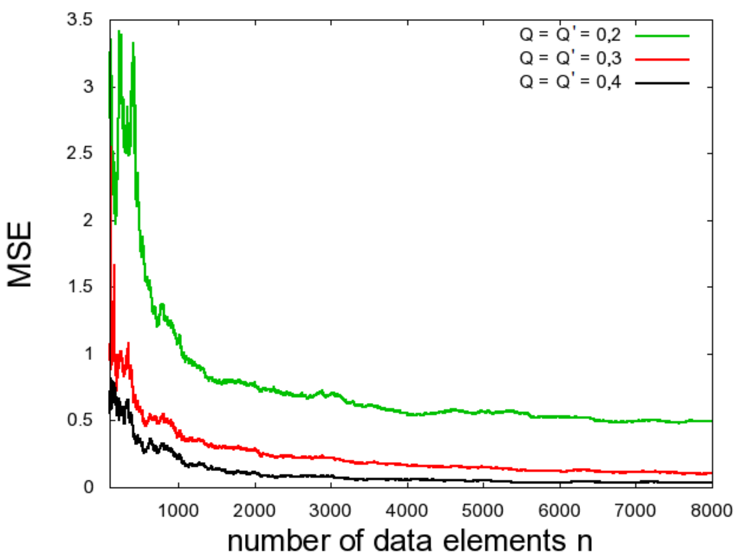

Fig. 3. MSE values for estimator $\hat{\phi}_n(x, \{\hat{a}_i^{(8)}\}_n)$ in a function of number of data elements n, for three different values of parameter $Q = Q'$: $Q = Q' = 0,2$, $Q = Q' = 0,3$ and $Q = Q' = 0,4$ ($alpha = 0,2$)

The constant a is estimated by making use of estimator (4), with parameter $\omega = 8$. The results of the simulation, obtained for $\alpha = 0,2$, are shown in Fig. 2.

The estimator $\hat{a}_n^{(8)}$ converges to value $2,5$ quite fast. This satisfactory result should be reflected in the quality of estimation of the regression function $\phi(x)$. To investigate the quality of estimator (31) the Mean Squared Error (MSE) value is calculated for each considered number of data elements n. Simulations are performed for three different values of parameter $Q = Q'$. Parameter α is set to $0,2$. Results are presented in Fig. 3.

For all considered values of $Q = Q'$ the estimator $\hat{\phi}_n(x, \{\hat{a}_i^{(8)}\}_n)$ seems to converge to the regression function $\phi(x)$ as the number of data elements n increases. It is easily seen that for $\alpha = 0,2$, inequalities (39) are satisfied. An interesting question arrised how the estimator (31) behaves if inequalities (39) are not held. To answer this question simulations for three different values of α are performed, keeping fixed $Q = Q' = 0,3$. According to inequalities (39), for $Q = 0,3$ the parameter α should obey $\alpha < 0,5$. Obtained results are shown in Fig. 4.

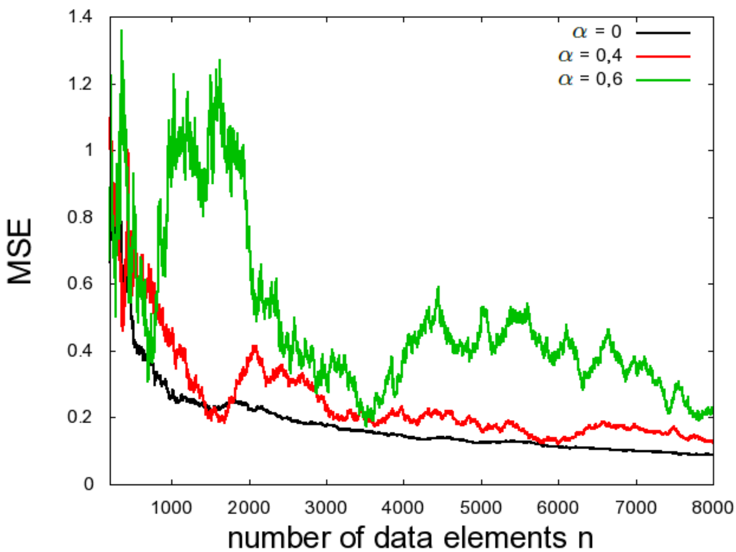

Fig. 4. MSE values for estimator $\hat{\phi}_n(x, \{\hat{a}_i^{(8)}\}_n)$ in a function of number of data elements n, for three different values of α: $\alpha = 0$, $\alpha = 0,4$ and $\alpha = 0,6$ ($Q = Q' = 0,3$)

In Figure 5 an example estimator $\hat{\phi}_n(x, \{\hat{a}_i^{(8)}\}_n)$, obtained for $n = 8000$, $Q = Q = 0,3$ and $\alpha = 0,2$ is presented in comparison with the regression function (32).

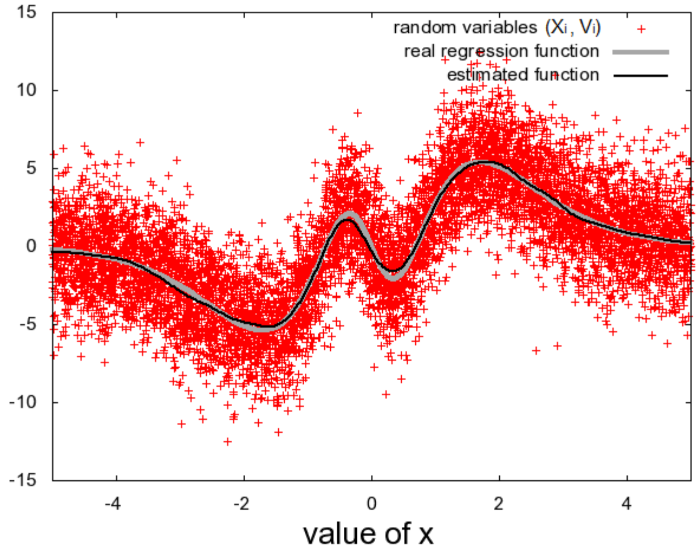

Fig. 5. Estimator $\hat{\phi}_n(x, \{\hat{a}_i^{(\omega)}\}_n)$, obtained for $\alpha = 0,2$, $Q = Q' = 0,3$ and $n = 8000$, in comparison with the regression function $\phi(x)$. Points denote the random variable pairs (X_i, V_i).

6 Conclusions and Future Work

In the paper the recursive least squares method, combined with the general regression neural networks, was presented. Thess tools were applied to solve the problem of learning in time-varying environment. The general regression neural network were developed using the orthogonal-type kernel functions. Future work can be focused on applying some other methods, e.g. supervised and unsupervised neural networks (see e.g. [2], [3] and [11]) or neurofuzzy structures (see e.g. [10], [31]-[34], [37] and [38]), to handle non-stationary noise. Moreover, the recursive form of the algorithm presented in this paper allows to adopt it for data streams.

Acknowledgments. The paper was prepared under project operated within the Foundation for Polish Science Team Programme co-financed by the EU European Regional Development Fund, Operational Program Innovative Economy 2007-2013, and also supported by National Science Centre NCN.

References

1. Albert, A.E., Gardner, L.A.: Stochastic Approximation and Nonlinear Regression, No. 42. MIT Press, Cambridge (1967)
2. Bilski, J., Rutkowski, L.: A fast training algorithm for neural networks. IEEE Transactions on Circuits and Systems II 45, 749–753 (1998)

3. Cierniak, R., Rutkowski, L.: On image compression by competitive neural networks and optimal linear predictors. Signal Processing: Image Communication - a Eurasip Journal 15(6), 559–565 (2000)
4. Gałkowski, T., Rutkowski, L.: Nonparametric recovery of multivariate functions with applications to system identification. Proceedings of the IEEE 73, 942–943 (1985)
5. Gałkowski, T., Rutkowski, L.: Nonparametric fitting of multivariable functions. IEEE Transactions on Automatic Control AC 31, 785–787 (1986)
6. Greblicki, W., Pawlak, M.: Nonparametric system indentification. Cambridge University Press (2008)
7. Greblicki, W., Rutkowska, D., Rutkowski, L.: An orthogonal series estimate of time-varying regression, Tokyo. Annals of the Institute of Statistical Mathematics, vol. 35, Part A, pp. 147–160 (1983)
8. Greblicki, W., Rutkowski, L.: Density-free Bayes risk consistency of nonparametric pattern recognition procedures. Proceedings of the IEEE 69(4), 482–483 (1981)
9. Györfi, L., Kohler, M., Krzyżak, A., Walk, H.: A Distribution-Free Theory of Nonparametric Regression. Springer Series in Statistics. Springer, USA (2002)
10. Nowicki, R.: Nonlinear modelling and classification based on the MICOG defuzzifications. Journal of Nonlinear Analysis, Series A: Theory, Methods and Applications 7(12), 1033–1047 (2009)
11. Patan, K., Patan, M.: Optimal Training Strategies for Locally Recurrent Neural Networks. Journal of Artificial Intelligence and Soft Computing Research 1(2), 103–114 (2011)
12. Rafajłowicz, E.: Nonparametric orthogonal series estimators of regression: A class attaining the optimal convergence rate in L_2. Statistics and Probability Letters 5, 219–224 (1987)
13. Rutkowski, L.: Sequential estimates of probability densities by orthogonal series and their application in pattern classification. IEEE Transactions on Systems, Man, and Cybernetics SMC 10(12), 918–920 (1980)
14. Rutkowski, L.: Sequential estimates of a regression function by orthogonal series with applications in discrimination, New York, Heidelberg, Berlin. Lectures Notes in Statistics, vol. 8, pp. 236–244 (1981)
15. Rutkowski, L.: On system identification by nonparametric function fitting. IEEE Transactions on Automatic Control AC 27, 225–227 (1982)
16. Rutkowski, L.: Orthogonal series estimates of a regression function with applications in system identification. In: Probability and Statistical Inference, pp. 343–347. D. Reidel Publishing Company, Dordrecht (1982)
17. Rutkowski, L.: On Bayes risk consistent pattern recognition procedures in a quasi-stationary environment. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI 4(1), 84–87 (1982)
18. Rutkowski, L.: On-line identification of time-varying systems by nonparametric techniques. IEEE Transactions on Automatic Control AC 27, 228–230 (1982)
19. Rutkowski, L.: On nonparametric identification with prediction of time-varying systems. IEEE Transactions on Automatic Control AC 29, 58–60 (1984)
20. Rutkowski, L.: Nonparametric identification of quasi-stationary systems. Systems and Control Letters 6, 33–35 (1985)
21. Rutkowski, L.: The real-time identification of time-varying systems by nonparametric algorithms based on the Parzen kernels. International Journal of Systems Science 16, 1123–1130 (1985)

22. Rutkowski, L.: A general approach for nonparametric fitting of functions and their derivatives with applications to linear circuits identification. IEEE Transactions Circuits Systems CAS 33, 812–818 (1986)
23. Rutkowski, L.: Sequential pattern recognition procedures derived from multiple Fourier series. Pattern Recognition Letters 8, 213–216 (1988)
24. Rutkowski, L.: Nonparametric procedures for identification and control of linear dynamic systems. In: Proceedings of 1988 American Control Conference, June 15-17, pp. 1325–1326 (1988)
25. Rutkowski, L.: An application of multiple Fourier series to identification of multivariable nonstationary systems. International Journal of Systems Science 20(10), 1993–2002 (1989)
26. Rutkowski, L.: Nonparametric learning algorithms in the time-varying environments. Signal Processing 18, 129–137 (1989)
27. Rutkowski, L., Rafajłowicz, E.: On global rate of convergence of some nonparametric identification procedures. IEEE Transaction on Automatic Control AC 34(10), 1089–1091 (1989)
28. Rutkowski, L.: Identification of MISO nonlinear regressions in the presence of a wide class of disturbances. IEEE Transactions on Information Theory IT 37, 214–216 (1991)
29. Rutkowski, L.: Multiple Fourier series procedures for extraction of nonlinear regressions from noisy data. IEEE Transactions on Signal Processing 41(10), 3062–3065 (1993)
30. Rutkowski, L., Gałkowski, T.: On pattern classification and system identification by probabilistic neural networks. Applied Mathematics and Computer Science 4(3), 413–422 (1994)
31. Rutkowski, L.: A New Method for System Modelling and Pattern Classification. Bulletin of the Polish Academy of Sciences 52(1), 11–24 (2004)
32. Rutkowski, L., Cpałka, K.: A general approach to neuro-fuzzy systems. In: Proceedings of the 10th IEEE International Conference on Fuzzy Systems, Melbourne, December 2-5, vol. 3, pp. 1428–1431 (2001)
33. Rutkowski, L., Cpałka, K.: A neuro-fuzzy controller with a compromise fuzzy reasoning. Control and Cybernetics 31(2), 297–308 (2002)
34. Scherer, R.: Boosting Ensemble of Relational Neuro-fuzzy Systems. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Żurada, J.M. (eds.) ICAISC 2006. LNCS (LNAI), vol. 4029, pp. 306–313. Springer, Heidelberg (2006)
35. Specht, D.F.: Probabilistic neural networks. Neural Networks 3, 109–118 (1990)
36. Specht, D.F.: A general regression neural network. IEEE Transactions on Neural Networks 2, 568–576 (1991)
37. Starczewski, J., Rutkowski, L.: Interval type 2 neuro-fuzzy systems based on interval consequents. In: Rutkowski, L., Kacprzyk, J. (eds.) Neural Networks and Soft Computing, pp. 570–577. Physica-Verlag, Springer-Verlag Company, Heidelberg, New York (2003)
38. Starczewski, J.T., Rutkowski, L.: Connectionist Structures of Type 2 Fuzzy Inference Systems. In: Wyrzykowski, R., Dongarra, J., Paprzycki, M., Waśniewski, J. (eds.) PPAM 2001. LNCS, vol. 2328, pp. 634–642. Springer, Heidelberg (2002)
39. Szegö, G.: Orthogonal Polynomials, vol. 23. American Mathematical Society Coll. Publ. (1959)
40. Wilks, S.S.: Mathematical Statistics. John Wiley, New York (1962)