Marcin PLUCIŃSKI

West Pomeranian University of Technology, Faculty of Computer Science and Information Technology

Nonlinear ellipsoidal mini-models – application for the function approximation task

Abstract. *Mini-models are local regression models, which can be used for the function approximation learning. In the paper, there are presented mini-models based on hyper-spheres and hyper-ellipsoids and researches were made for linear and nonlinear models with no limitations for the problem input space dimension. Learning of the approximation function based on mini-models is very fast and it proved to have a good accuracy. Mini-models have also very advantageous extrapolation properties.*

Streszczenie. *Mini-modele to modele lokalnej regresji, które można wykorzystać do aproksymacji funkcji. W artykule opisano mini-modele o bazie hiper-sferycznej i hiper-elipsoidalnej oraz badania dla mini-modeli linowych i nieliniowych bez ograniczeń na rozmiar przestrzeni wejść. Uczenie aproksymującej funkcji opartej na mini-modelach jest szybkie, a sama funkcja ma dobrą dokładność i korzystne własności ekstrapolacyjne. (**Nieliniowe, elipsoidalne mini-modele – zastosowanie do aproksymacji funkcji**).*

Keywords: mini-model, local regression, k-nearest neighbours method, function approximation.
Słowa kluczowe: mini-modele, regresja lokalna, metoda k-najbliższych sąsiadów, aproksymacja funkcji.

Introduction

Learning of function approximators with an application of so called memory-based learning methods is very often attractive approach in comparison with creating of global models based on a parametric representation. In some situations (for example: small number of samples), building of global models can be difficult and then memory-based methods become one of possible solutions for the approximation task.

Memory-based methods are very well explored and described in many bibliography positions. The most important here is the k nearest neighbours method (kNN), which is described in many versions [1,2,3], but still is the subject of new researches [4,5]. Another approaches can be methods based on locally weighted learning [1,6] which use different ways of a samples weighting. Methods widely applied in this category are, for example, probabilistic neural networks and generalised regression networks [7,8].

The concept of mini-models was introduced by prof. Andrzej Piegat. In papers [9,10] there were described local regression models based on simplexes. Described models were linear and a research work was made only for problems in a 1 and 2-dimensional input space. This paper presents mini-models based on hyper-spheres and hyper-ellipsoids. Researches were made for linear and nonlinear models with no limitations for the problem input space dimension.

Mini-models with a hyper-spherical base

The main idea of mini-models is similar to the kNN method. During calculations of an answer for a question point $\mathbf{x}^*$ only k nearest (in a meaning of an applied metric – here Euclidean metric) samples are taken into account. In the classic kNN method the model answer is calculated as a mean value of target function values or a weighted mean value. In such case, weight values usually depend on a distance $\delta(\mathbf{x}^*,\mathbf{x})$ between the question point $\mathbf{x}^*$ and analysed neighbours $\mathbf{x}$, for example:

(1) $$w_{x^*,x} = \frac{1}{1 + m \cdot \delta(\mathbf{x}^*, \mathbf{x}) / k^2},$$

where: the m parameter is taken empirically.

The mini-model is a local regression and the answer for the question point $\mathbf{x}^*$ is calculated on the base of a local model created for k nearest neighbours. The mini-model is always created in time of answer calculations.

In the simplest case the linear mini-model can be applied and then the answer is calculated on the base of the linear regression:

(2) $$f(\mathbf{x}^*) = \mathbf{w}^{\mathrm{T}} \cdot \mathbf{x}^*,$$

where: $\mathbf{w}$ – the vector of linear mini-model coefficients found for k-neighbours.

In papers [9,10] there are described mini-models created for sectors of the input space that have a triangle shape (in a 2-dimensional input space) or a simplex shape in a multi-dimensional input space. Such sector will be called a mini-model base. In this paper, mini-models will be created for a circular base in a 2-dimensional input space or a spherical (hiper-spherical) base in a 3 or multi-dimensional input space.

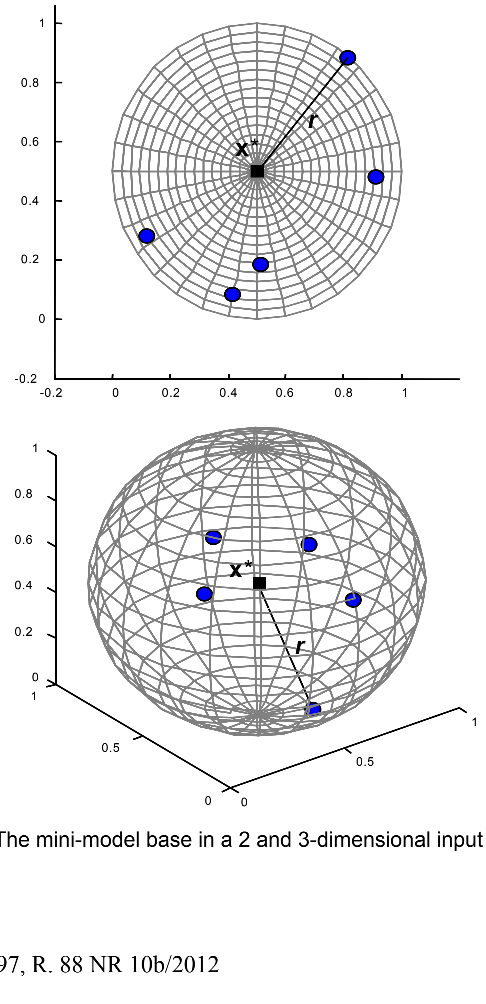

Fig.1. The mini-model base in a 2 and 3-dimensional input space

The mini-model base has a center in the question point $\mathbf{x}^*$ and its radius r is defined by a distance between the point $\mathbf{x}^*$ and the most distant point from k neighbours, Fig. 1.

Nonlinear mini-models have better possibilities of fitting to the samples. An answer of such model is a sum of a linear mini-model and an additional nonlinear component:

(3) $$f(\mathbf{x}^*) = \mathbf{w}^{\mathrm{T}} \cdot \mathbf{x}^* + f_N(\mathbf{x}^*).$$

As the mini-model is usually created for a small number k of nearest neighbours, the nonlinear function f_N should have a possibility of changing its shape thanks to as small number of coefficients as possible (because n+1 coefficients must be tuned in the vector $\mathbf{w}$).

Among many inspected functions, very advantageous properties has the function:

(4) $$f_N(\mathbf{x}) = w_N \cdot \sin\left[\frac{\pi}{2} - \left\|\mathbf{x} - \mathbf{x}^*\right\| \frac{\pi}{r}\right],$$

where: r is the radius of the mini-model base. In such created function we have only one coefficient w_N to learn, Fig. 2. During learning we must find such a w_N value to obtain the best fit of the mini-model to k neighbours. Exemplary linear and nonlinear mini-models in a 1 and 2-dimensional input space are presented in Fig. 3 and Fig. 4.

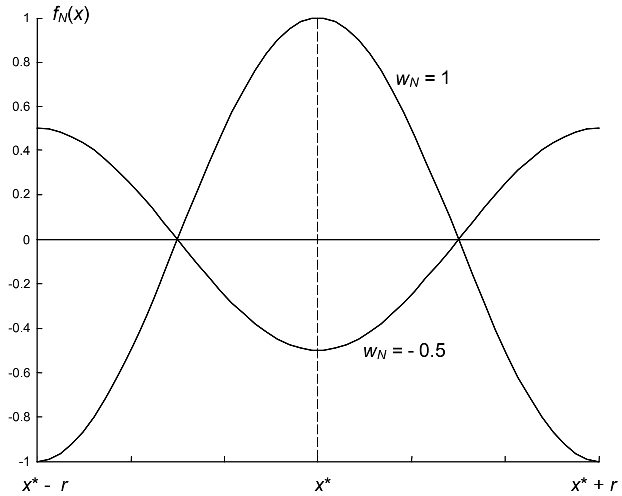

Fig.2. Exemplary shapes of the mini-model nonlinear component for w_N = 1 and w_N = –0.5

Mini-models with a hyper-ellipsoidal base

Very often data are not placed uniformly in the input space. Such situation usually happens in the case of multi-dimensional input space. For such data, a quality of modelling could be improved by application of mini-models with the base different than hyper-spherical. Mini-models with the simplex base, proposed in [9,10], are more flexible in fitting to the data, but it is difficult to apply them in the multi-dimensional input space.

A mini-model which is more flexible in fitting to the data, and which is still easy to tune, is the mini-model with a hyper-ellipsoidal base. An example of such mini-model base in a 2-dimensional input space is presented in Fig. 5.

During tuning of mini-model parameters, the data that are chosen to create mini-model base (k nearest neighbours) are a subject of the PCA transformation and next they are normalised. The process of local data transformation is illustrated in Fig. 6.

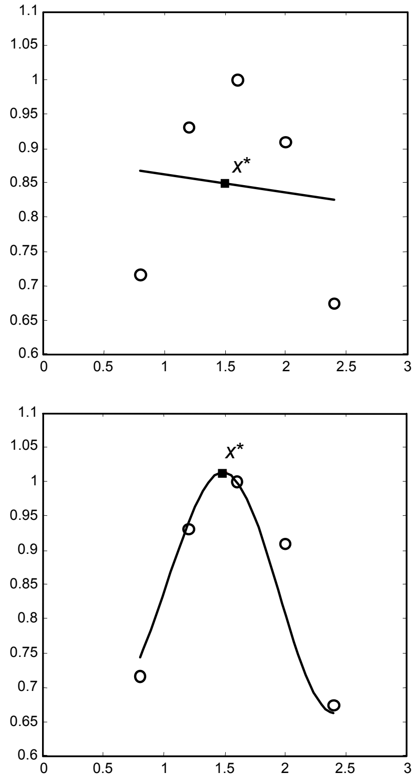

Fig.3. Exemplary linear and nonlinear mini-model in a 1-dimensional input space

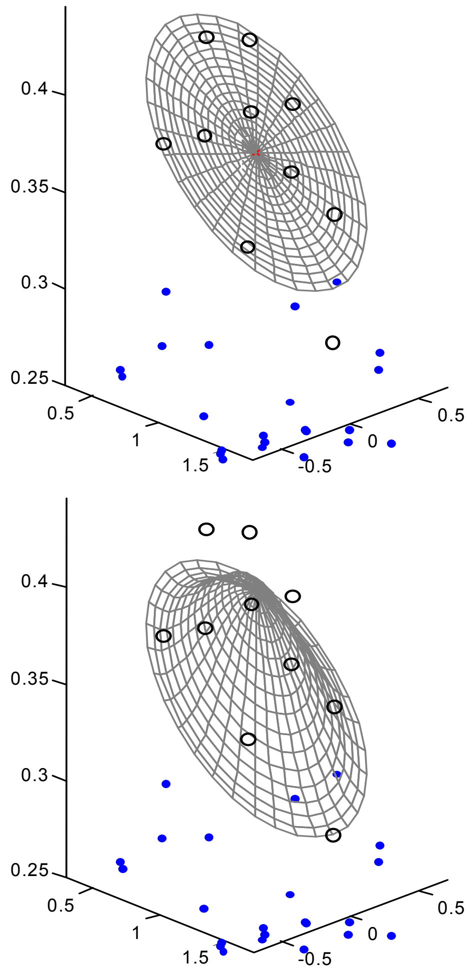

Fig.4. Exemplary linear and nonlinear mini-model in a 2-dimensional input space

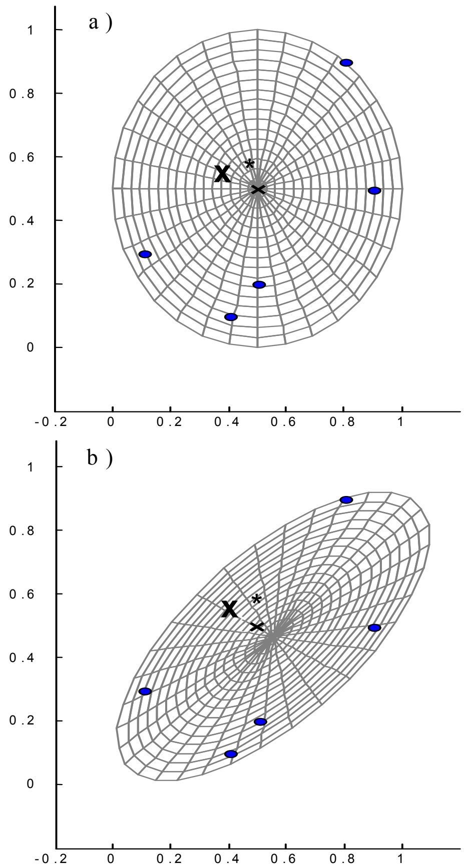

Fig.5. An example of circular and elliptic mini-model base in a 2-dimensional input space

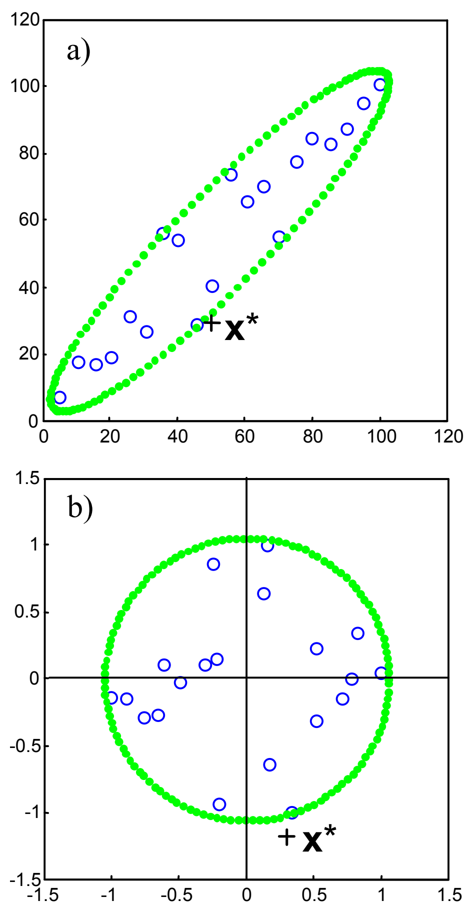

Fig.6. Exemplary local data before (a) and after (b) transformation

After transformation, samples are uniformly placed in the entire new input space, Fig. 6b, and their values belong to the normalised range [-1,1]. Such created new local data are used for creating nonlinear hyper-spherical mini-model on the base of equations similar to (3) and (4). The mini-model has a hyper-spherical base in the new transformed input space, but in the original input space its base has a hyper-ellipsoidal shape and in that way it can better fit into the data, Fig. 6a and 7. It must be emphasised here that mini-model with a more complex shape of the base is created with the same number of coefficients as hyper-spherical one and its tuning is not changed (only PCA transformation must be realised in the beginning, but it is fast because is performed on a small number of samples).

There is small difference in the calculation of the answer of the ellipsoidal mini-model. The mini-model base center doesn't lie now in the question point $\mathbf{x}^*$, but in the beginning of the new (after transformation) coordinate system. It is illustrated in Fig. 6.

The entire algorithm of calculating the answer for the question point $\mathbf{x}^*$ can be described in following steps.

1. Find k nearest neighbours of the question point $\mathbf{x}^*$.
2. Perform PCA transformation and normalisation of chosen samples.
3. Perform transformation of the question point $\mathbf{x}^*$ into the new coordinate system $\mathbf{x}^* \rightarrow \mathbf{x}^*_{PCA}$.
4. Tune the mini-model parameters and calculate its output for $\mathbf{x}^*_{PCA}$ according to formulas:

(5) $$f(\mathbf{x}^*_{PCA}) = \mathbf{w}^{\mathrm{T}} \cdot \mathbf{x}^*_{PCA} + f_N(\mathbf{x}^*_{PCA}),$$

(6) $$f_N(\mathbf{x}_{PCA}) = w_N \cdot \sin\left[\frac{\pi}{2} - \|\mathbf{x}_{PCA}\| \cdot \frac{\pi}{r}\right],$$

where: r = 1 in the new coordinate system (after transformation and normalisation).

An example of the ellipsoidal mini-model created for samples in a 2-dimensional input space is presented in Fig. 7.

Experiments

For better visualisation of mini-models work, first experiments were realised for data with 1 and 2-dimensional input space. Of course, spherical and ellipsoidal mini-models are the same in a 1-dimensional input space. Fig. 8 presents the characteristic of the model created with an application of nonlinear mini-models. For comparison, there is also presented the characteristic of the model created by the kNN method.

Mini-models have a very good extrapolation property what is presented in Fig. 9. As before, there are presented characteristics of models created with an application of the kNN method and mini-models. First of all, an attention should be paid for a behaviour of mini-models in places where there are no samples (information gaps) and outside of the samples input domain. Mini-models give answers that are much more consistent with a common sense and a shape of their characteristic lines is smoother.

Fig. 10 presents surfaces created for samples with a 2-dimensional input space. Both mini-models approximators from Fig. 8 and Fig. 10 have better accuracy than the kNN method and values of model real errors are given in the Table 1.

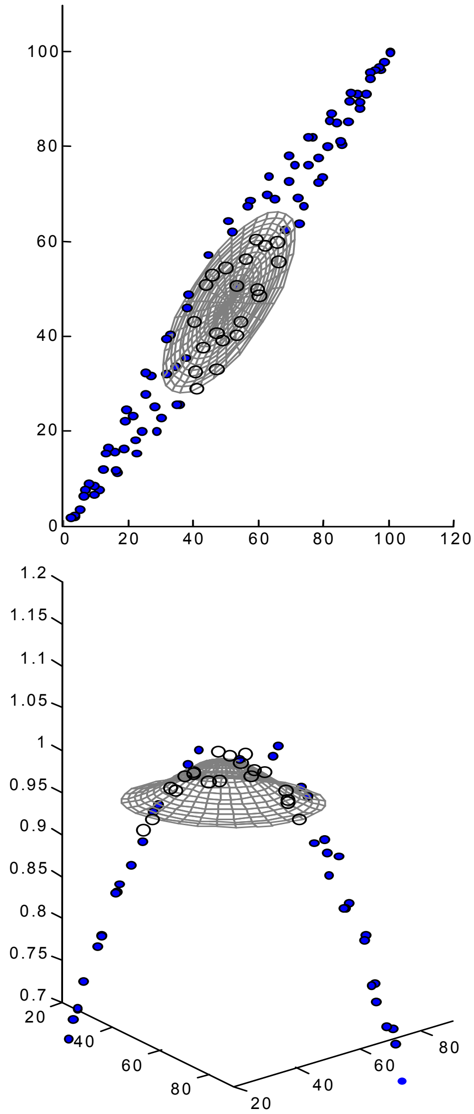

Fig.7. An example of ellipsoidal mini-model created for samples in a 2-dimensional input space

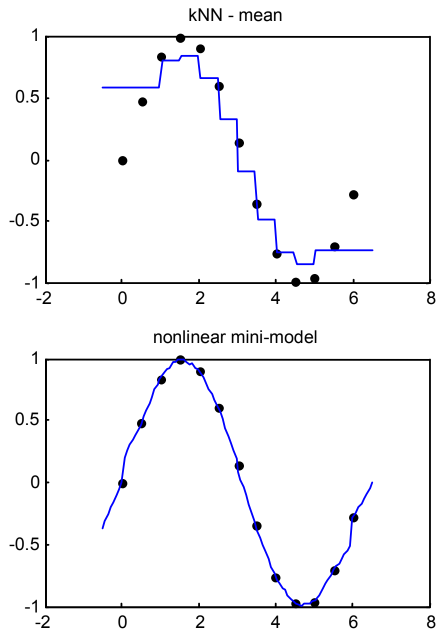

Fig.8. Characteristics of approximators created for data with a 1-dimensional input space by the *k*NN method and nonlinear mini-models

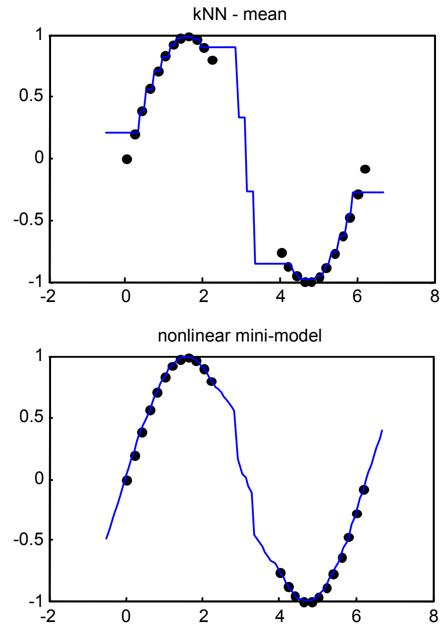

Fig.9. Characteristics of models created for data with the information gap

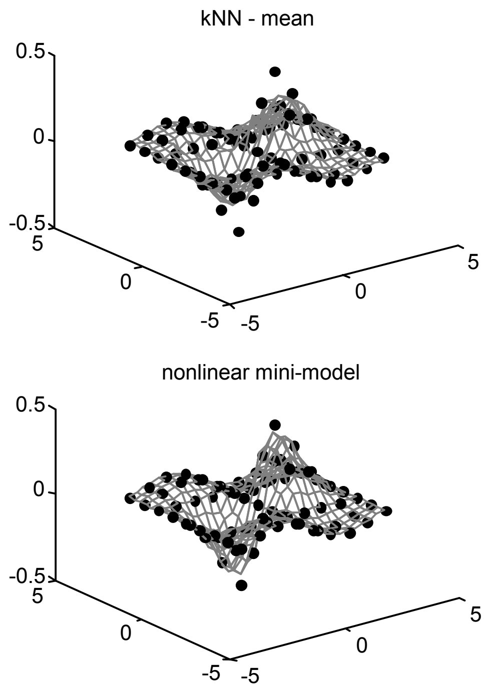

Fig.10. Characteristics of approximators created for data with a 2-dimensional input space by the *k*NN method and nonlinear mini-models

In the next part of experiments, there was evaluated a real accuracy of function approximators based on mini-models, Table 1. The research work was performed on data created by the author and data from popular web repositories. Data were normalised due to different ranges of its inputs.

The real error was calculated with an application of the leave one out crossvalidation method. Mini-models approximators are compared with *k*NN method and in each case calculation results are presented for an optimal number of neighbours (giving the lowest real error). Additionally, for comparison purpose, there is also given an approximation accuracy for a generalised regression network (GRN) also with a neuron width optimally tuned.

Table 1. The real error of function approximators

data	inputs number	mean kNN	weighted mean kNN	linear mini-model	nonlinear hyper-spherical mini-model	nonlinear hyper-ellipsoidal mini-model	GRN
$\sin(x)$ with 0.5 sampling step (Fig. 8)	1	0.166	0.161	0.095	0.086	0.086	0.142
$\sin(x)$ with 0.2 sampling step	1	0.031	0.030	0.013	0.008	0.008	0.025
$\frac{\sin(x_1)\cdot\sin(x_2)}{x_1^2+x_2^2+1}$ with 0.25 samp. step (Fig. 10)	2	0.0281	0.0277	0.0228	0.0259	0.0237	0.0254
$\frac{\sin(x_1)\cdot\sin(x_2)}{x_1^2+x_2^2+1}$ with 0.1 sampling step	2	0.0058	0.0058	0.0049	0.0055	0.0054	0.0052
`bodyfat`	14	2.236	2.211	0.472	0.475	0.476	2.671
`cpu`	6	29.507	28.937	27.305	27.826	26.779	28.771
`diabetes_numeric`	2	0.474	0.481	0.475	0.487	0.479	0.471
`housing`	13	2.771	2.685	2.289	2.259	2.346	2.506
`elusage`	2	8.885	8.984	8.648	9.154	8.539	9.323

Interesting results were obtained for "`elusage`" data. Mini-models with the ellipsoidal base gave in this case the smallest real error. The input space of this data is 2-dimensional so it is possible to visualise it, Fig. 11. It can be seen that the distribution of samples is not uniform – data create a path. In such situations mini-models with the hyper-ellipsoidal base have a chance to work with better accuracy than mini-models with hyper-spherical base.

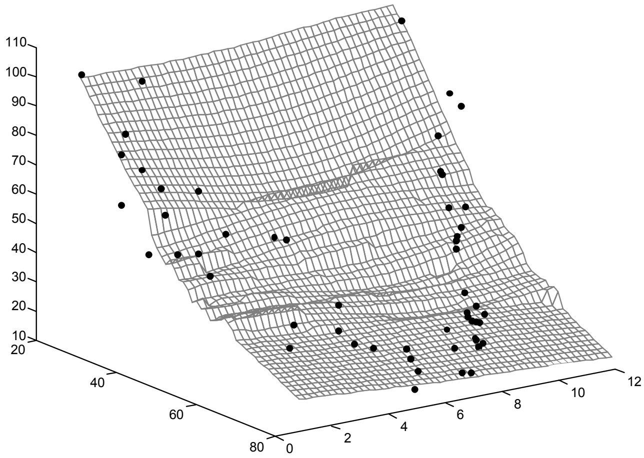

Fig.11. Distribution of samples in "`elusage`" data and the surface of the model characteristic created by elliptic mini-models

Conclusions

First of all, the approximation function based on mini-models proved to have a good accuracy, Table 1. The accuracy is particularly great for data without noise. In the case of noised data, mini-models have the accuracy comparable or slightly worse than kNN methods.

Learning of the approximator is very fast – it is enough to memorise learning data and the proper mini-model is created only in time of calculating an answer for the question point $\mathbf{x}^*$. A mini-models creating is not computationally complex because they are build on the base of a small number of samples (k nearest neighbours). The linear mini-model is a linear regression found for k neighbours and the nonlinear one has only one additional coefficient to compute.

Mini-models have very advantageous extrapolation properties. It results from a fact, that they take into account not only samples target values, but also a tendency in the neighbourhood of the question point. Using information about this tendency cause better modelling in places where there is no data (information gaps and outside of the input space domain). Information gaps are very characteristic property of multi-dimensional data with a small number of samples. In such situation data are not placed uniformly in the input space and it is advantageous to apply mini-models with the hyper-ellipsoidal base.

A weakness of mini-models in comparison with the kNN method is a necessity of taking into account a greater number of neighbours k. For example, a minimal number of neighbours needed for linear mini-model is equal n+1, where n is a size of an input space.

Wydanie publikacji zrealizowano przy udziale środków finansowych otrzymanych z budżetu Województwa Zachodniopomorskiego.

REFERENCES

[1] Cichosz P., Learning systems. WNT Publishing House, Warsaw, 2000, [in Polish]

[2] Hand D., Mannila H., Smyth P., Principles of data mining. The MIT Press, 2001

[3] Moore A.W., Atkeson C.G., Schaal S.A., Memory-based learning for control. *Technical Report CMU-RI-TR-95-18*, Carnegie-Mellon University, Robotics Institute, 1995

[4] Kordos M., Blachnik M., Strzempa D., Do we need whatever more than k-NN? *Proceedings of 10-th International Conference on Artificial Intelligence and Soft Computing*, Zakopane, Poland, pp. 414-421, Springer, 2010

[5] Korzeń M., Klęsk P., Sets of approximating functions with finite Vapnik-Czervonenkis dimension for nearest-neighbours algorithm. *Pattern Recognition Letters*, 32, pp. 1882-1893, 2011

[6] Atkeson C.G., Moore A.W., Schaal S.A., Locally weighted learning. *Artificial Intelligence Review*, 11, pp. 11-73, 1997

[7] Pluciński M., Application of data with missing attributes in the probability RBF neural network learning and classification. *Artificial Intelligence and Security in Computing Systems: 9th International Conference ACS'2002: Proceedings*, Eds.: J. Sołdek, L. Drobiazgiewicz, Boston/Dordrecht/London: Kluwer Academic Publishers, pp. 63-72, 2003

[8] Wasserman P.D., Advanced methods in neural computing. New York, Van Nostrand Reinhold, 1993

[9] Piegat A., Wąsikowska B., Korzeń M., Application of the self-learning, 3-point mini-model for modelling of unemployment rate in Poland. *Studia Informatica*, University of Szczecin, 2010, [in Polish]

[10] Piegat A., Wąsikowska B., Korzeń M., Differences between the method of mini-models and the k-nearest neighbours on example of modelling of unemployment rate in Poland. *Proceedings of 9th Conference on Information Systems in Management*, pp. 34-43, WULS Press, Warsaw, 2011

***Author**: dr inż. Marcin Pluciński, West Pomeranian University of Technology, Faculty of Computer Science and Information Technology, Żołnierska 49, 71-210 Szczecin, Poland, E-mail: mplucinski@wi.zut.edu.pl*