Computer Assisted Methods in Engineering and Science, **21**: 77–84, 2014.

Application of support vector machine in geodesy for the classification of vertical displacements

Maria Mrówczyńska
University of Zielona Góra, Szafrana 1, 65-516 Zielona Góra, Poland
e-mail: M.Mrowczynska@ib.uz.zgora.pl

The article presents basic rules for constructing and training neural networks, called the Support Vector Machine technique. SVM networks can mainly be used for solving tasks of classification of linearly and nonlinearly separable data and regression as well as identifying signals and recognising increases.

In this paper SVM networks have been used for classifying linearly separable data in order to formulate a model of displacements of points representing a monitored object. The problem of learning networks requires the use of quadratic programming in search of an optimum point of a Lagrange function with respect to optimised parameters. Estimated parameters determine the location of the hyperplane which maximises the separation margin of both classes.

Keywords: linear SVM network, classification, displacements.

1. Introduction

SVM classifiers, called the SVM technique, were invented by Vapnik [10]. They are a new approach to constructing and training networks. SVM networks belong to a group of unidirectional networks, they usually have a two-layer structure and can use different types of activation function [7, 11]. In comparison with MLP (Multi Layer Perceptron) networks, SVM networks do not have the faults typical of MLP networks, i.e. the possibility that the minimisation process stops at one of many local minima, or an initially adopted arbitrary architecture of the network, which determines its future generalisation abilities.

Generally speaking, in the case of linearly separable data, SVM networks make it possible to find an optimum separation hyperplane which separates classes with a possibly maximum separation margin (Fig. 1). In the case of linearly inseparable data, we find a curvilinear classification boundary. A separation margin is the distance between an optimum hyperplane and the closest vector. The paper presents the problem of the use of an SVM network for the classification of linearly separable data, which will be illustrated by measurements of height differences between points in a measurement and control network. The measurement and control network has been set up on a building located on expansive soil. In addition, a case of the measurement of deviations in the horizontal and vertical plane of components of civil engineering structures will be discussed.

2. Linear SVM network in a classification task

If we assume that a set of training pairs $(\mathbf{x}_i, d_i)$, $i = 1, \ldots, p$ undergoes classification, where $\mathbf{x}_i$ is an input vector, and the setpoint d_i equals 1 or -1, then the equation of a hyperplane separating the two classes will be written as

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0, \tag{1}$$

where N-dimensional weight vector $\mathbf{w} = [w_1, w_2, \ldots, w_N]$, the input vector $\mathbf{x} = [x_1, x_2, \ldots, x_N]$, and the weight b is polarisation.

The equation that enables the determination of a point's membership in a particular class takes the form

$$\begin{aligned} &\mathbf{w}^T\mathbf{x}_i + b > 0 \rightarrow d_i = 1, \\ &\mathbf{w}^T\mathbf{x}_i + b < 0 \rightarrow d_i = -1. \end{aligned} \tag{2}$$

An optimum hyperplane $g(\mathbf{x}) = \mathbf{w}_o^T\mathbf{x} + b_o = 0$ is one for which the separation margin is maximum (Fig. 1).

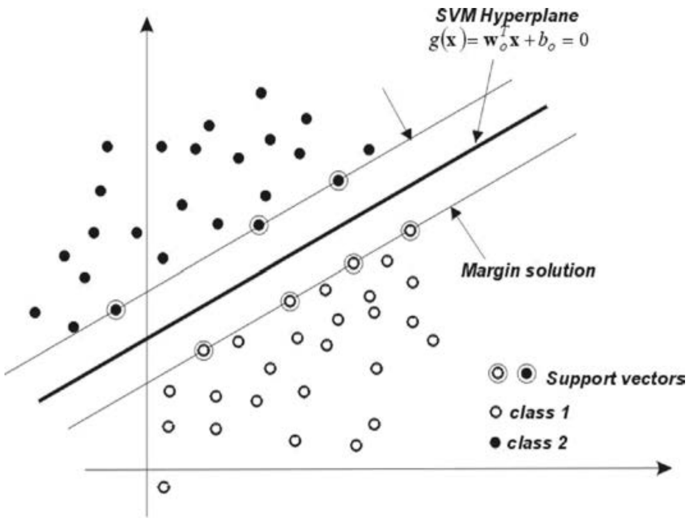

Fig. 1. Optimum hyperplane with a maximum separation margin.

In order to define and determine the separation margin it is necessary to find the distance between any point $\mathbf{x}$ and an optimum hyperplane. This distance is expressed by formula

$$r(\mathbf{x}) = \frac{g(x)}{\|\mathbf{w}_o\|}. \tag{3}$$

After the normalisation of the Eqs. (2), the decisive equations determining the membership of data in a particular class take the form

$$\begin{aligned} &\mathbf{w}^T\mathbf{x}_i + b \geq 1 \rightarrow d_i = 1, \\ &\mathbf{w}^T\mathbf{x}_i + b \leq -1 \rightarrow d_i = -1. \end{aligned} \tag{4}$$

If a pair of points $(\mathbf{x}_i, d_i)$ satisfies Eq. (4) with the equation mark, then a supporting vector $\mathbf{x}_{sv}$ is created, the distance between which and the hyperplane is expressed by the formula [1, 7]

$$\tau = \frac{g\left(\mathbf{x}_{sv}\right)}{\|\mathbf{x}\|} = \begin{cases} \dfrac{1}{\|\mathbf{w}\|} & \text{for} \quad g\left(\mathbf{x}_{sv}\right) = 1, \\[2ex] -\dfrac{1}{\|\mathbf{w}\|} & \text{for} \quad g\left(\mathbf{x}_{sv}\right) = -1. \end{cases} \tag{5}$$

The separation margin between two classes k will be designated as $k = 2\tau$. It is worth emphasising that supporting vectors are points located closest to the hyperplane, and for this reason they are the most difficult to classify, but they determine the location of the hyperplane and the width of the separation margin [2, 9].

3. Training an SVM network

Training an SVM network consists in choosing its weights in such a way as to have a maximum separation margin for linearly separable data, which satisfies condition (4). This kind of problem is called a primary problem, which will be written as

$$\min\left\{\frac{1}{2}\mathbf{w}^T\mathbf{w}\right\} \tag{6}$$

with the restrictions

$$d_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1. \tag{7}$$

The primary (optimisation) problem is a problem of quadratic programming with linear restrictions on weights. This kind of problem may be solved by means of Lagrange multipliers through the minimisation of the Lagrange function [1, 4]

$$J(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{p}\alpha_i\left[d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1\right]. \tag{8}$$

The subsequent multipliers α_i $i = 1, \ldots, p$, whose values correspond to particular restrictions, are the coordinates of the vector of Lagrange multipliers $\boldsymbol{\alpha}$. The minimisation of function (8) with respect to the optimised parameters $(\mathbf{w}, b, \boldsymbol{\alpha})$ consists in finding a saddle point of the function $J(\mathbf{w}, b, \boldsymbol{\alpha})$, assuming that it has been minimised with respect to vector $\mathbf{w}$ and parameter b, and maximises with respect to all the values of multipliers α_i. The conditions for optimality with respect to $\mathbf{w}$ and b can be written as

$$\frac{\partial J(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} - \sum_{i=1}^{p}\alpha_i d_i \mathbf{x}_i = 0, \tag{9}$$

$$\frac{\partial J(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial b} = 0 \rightarrow \sum_{i=1}^{p}\alpha_i d_i = 0, \tag{10}$$

which leads to a solution in the form

$$\mathbf{w} = \sum_{i=1}^{p}\alpha_i d_i \mathbf{x}_i. \tag{11}$$

While analysing formula (11) one can say that vector $\mathbf{w}$ is a linear combination of supporting vectors. Parameter b is determined by putting dependencies (9), (10) and (11) into formula (8), and a solution at the saddle point of the Lagrangean is obtained [13]

$$\mathbf{w}^T\mathbf{x}_{sv} + b = 1 \rightarrow b = 1 - \mathbf{w}^T\mathbf{x}_{sv}. \tag{12}$$

By putting dependence (11) into the basic form of the Lagrange function (8), the values of the unknowns in the form of vector $\mathbf{w}$ and parameter b are eliminated. Thus, the task of learning the network consists in maximising the Lagrange function with respect to multipliers α_i, and the primary problem is converted into a dual problem, which can be written as follows:

$$\max\left\{\sum_{i=1}^{p}\alpha_i - \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}\alpha_i\alpha_j d_i d_j \mathbf{x}_i^T\mathbf{x}_j\right\} \tag{13}$$

with the restrictions

$$\alpha_i \geq 0 \qquad \text{and} \qquad \sum_{i=1}^{p} \alpha_i d_i = 0. \tag{14}$$

In typical practical applications most Lagrange multipliers α_i equal zero. The data points $(\mathbf{x}_i, d_i)$, for which the corresponding Lagrange multipliers satisfy the condition $\alpha_i > 0$, will constitute supporting vectors $\mathbf{x}_{sv}$.

Example 1. *Let us analyse the procedure for determining a hyperplane and supporting vectors for the data in Table 1.*

Table 1. Data points for the illustrative example.

X		Y
1.25	2.33	1
1.75	0.55	1
3.82	3.07	1
4.85	0.25	−1
5.76	1.78	−1

The formulation of the problem of data classification by means of an SVM network consists in solving a quadratic programming task with respect to Lagrange multipliers (the minimisation of function (8)). In the first step a matrix $\mathbf{H}$ will be found, whose components (i, j) will be defined as

$$\mathbf{H}(i,j) = y_i y_j \mathbf{x}_i^T \mathbf{x}_j, \tag{15}$$

where

$$y_i y_j = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 \end{bmatrix},$$

$$\mathbf{x}_i^T \mathbf{x}_j = \begin{bmatrix} 6.9914 & 3.4690 & 11.9281 & 6.6450 & 11.3474 \\ 3.4690 & 3.3650 & 8.3735 & 8.6250 & 11.0590 \\ 11.9281 & 8.3735 & 24.0173 & 19.2945 & 27.4678 \\ 6.6450 & 8.6250 & 19.2945 & 23.5850 & 28.3810 \\ 11.3474 & 11.0590 & 27.4678 & 29.3810 & 36.3460 \end{bmatrix}$$

i.e.

$$\mathbf{H} = \begin{bmatrix} 6.9914 & 3.4690 & 11.9281 & -6.6450 & -11.3474 \\ 3.4690 & 3.3650 & 8.3735 & -8.6250 & -11.0590 \\ 11.9281 & 8.3735 & 24.0173 & -19.2945 & -27.4678 \\ -6.6450 & -8.6250 & -19.2945 & 23.5850 & 28.3810 \\ -11.3474 & -11.0590 & -27.4678 & 29.3810 & 36.3460 \end{bmatrix}.$$

Next, by means of quadratic programming, the equation of the hyperplane specified by the optimum weight vector

$$\mathbf{w}^T = \begin{bmatrix} -0.7490 & 0.4753 & 2.2715 \end{bmatrix}$$

and the parameter $b = 1$ will be determined. The values of the Lagrange multipliers are as follows:

$$\alpha^T = \begin{bmatrix} 0.0000 & 0.0000 & 0.3685 & 0.0000 & 0.3685 \end{bmatrix}.$$

The Lagrange multipliers α_i will be used to determine supporting vectors, which will be created exactly at the points for which $\alpha_i \neq 0$.

In the example above the width of the separation margin k equals

$$k = 2/\left\|\mathbf{w}\right\| = 2/\left\|0.7370\right\| = 2.71.$$

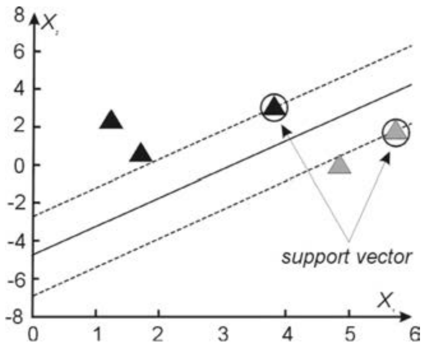

Fig. 2. Graphical representation of illustrative example.

4. Linearly inseparable data

The task of linearly inseparable data classification is an extremely difficult problem. When there is linear inseparability, the idea behind the SVM method is to construct an optimum hyperplane not in the input space but in a certain multidimensional space of the characteristics [9]. Therefore, the linearly inseparable data from an N – dimensional space (original space) will be transformed into a space of characteristics, which is a K – dimensional space, and $K > N$. While transforming the data from the original space to a space of characteristics the dimensionality is raised. The problem of data transformation has been shown in Fig. 3.

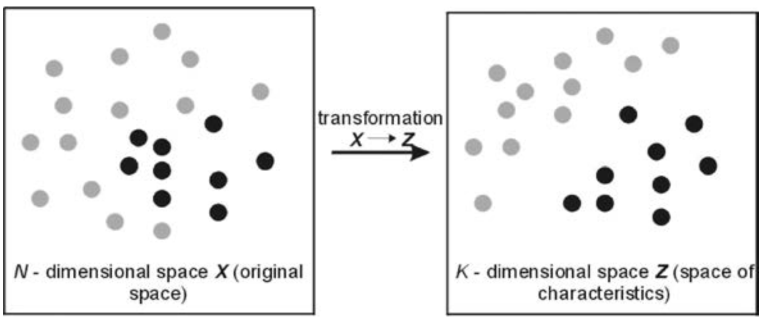

Fig. 3. Transformation of linearly inseparable data.

It is worth emphasising that raising the dimensionality allows one to find a separating hyperplane located in the space of characteristics whose picture can be seen in the original space. The equation of a hyperplane in the space of characteristics is written as the following formula

$$g(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i d_i K(\mathbf{x}, \mathbf{x}_i) + b, \tag{16}$$

where $K(\mathbf{x}, \mathbf{x}_i)$ is a scalar symmetric function called a kernel function. The kernel functions used most frequently are linear functions, polynomial functions, and radial and sigmoidal functions [8]. Depending on the kernel function used, there can be a slightly different data classification boundary. Examples of the kernel functions are shown in Table 2.

Table 2. Examples of kernel function.

Kernel type	Equation $K(\mathbf{x},\ \mathbf{x}_i)$	Comment
Linear	$K(\mathbf{x}, \mathbf{x}_i) = \mathbf{x}^T\mathbf{x}$	
Polynomial	$K(\mathbf{x}, \mathbf{x}_i) = \left(\mathbf{x}^T\mathbf{x} + 1\right)^b$	b – degree of the polynomial
Radial	$K(\mathbf{x}, \mathbf{x}_i) = \exp\left[-\left(1/2\sigma^2\right)\left(\lVert\mathbf{x} - \mathbf{x}_i\rVert^2\right)\right]$	σ – common to all kernel
Sigmoidal	$K(\mathbf{x}, \mathbf{x}_i) = \mathbf{x}^T\mathbf{x}$	constrains to β_0 and β_1

5. Numerical example

Civil engineering structures located on expansive soils are subjected mainly to vertical forces. In this case the problem of geodetic measurements focuses on examining the foundation settlement [3]. The object studied is represented by eleven controlled points stabilised on the building's walls at a height of about 0.50 m above the ground level. The layout of the measurement points was determined by the geological and engineering factors [6]. Each periodical measurement provided information on the changes in the behaviour of the object in the form of twenty-two observations carried out with the same structure of the measurement and control network. A principle had been adopted that measurement accuracy for relative displacements should not be less than ± 0.3 mm.

Using the notion of the reference system proposed by the author a model of displacements of controlled points has been formulated. The model a basis for making a two value decision about the linear separability of two classes of points with different directions of displacements. At the same time, it is necessary to avoid introducing decisions such as "reject", because it is not a two value decision if a particular point does not belong to either class. Now that the information about membership of the points in one of the two classes is available, the problem of the classifier consists in searching for a hyperplane separating the two classes with a maximum separation margin. As a result of the use of the SVM method, three supporting vectors have been determined, created at points with numbers 3, 7, 8, for which Lagrange coefficients are different from zero and the classifier forms a separation margin of about 2 m. The separation margin may help to make the right decision about the method of protecting a building subjected to uneven settlement.

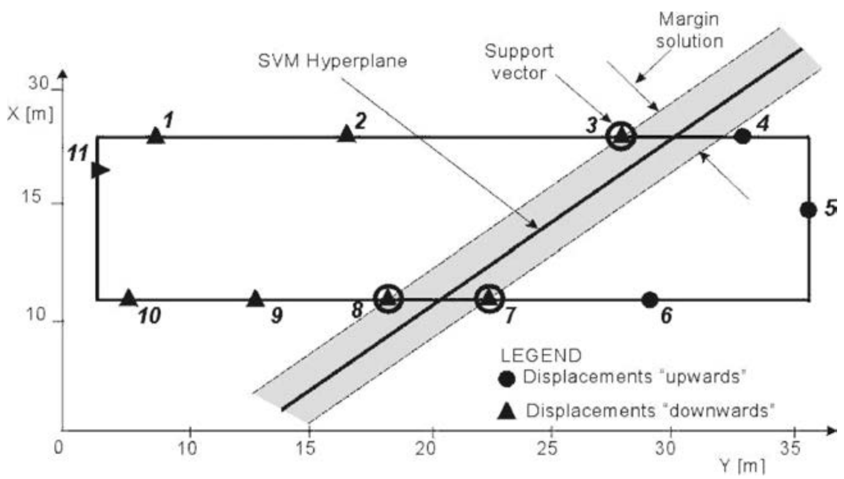

Fig. 4. Classification of points in a measurement and control network with SVM method.

In order to compare the results obtained with the SVM method, a geometrical displacement model has been created with the classic method. It makes use of an algorithm for the minimisation of absolute deviations and a criterion for a critical value of the increment of the square of the norm for the vector of corrections to the observations [3]. A graphic interpretation of the displacement values has been presented in Fig. 5.

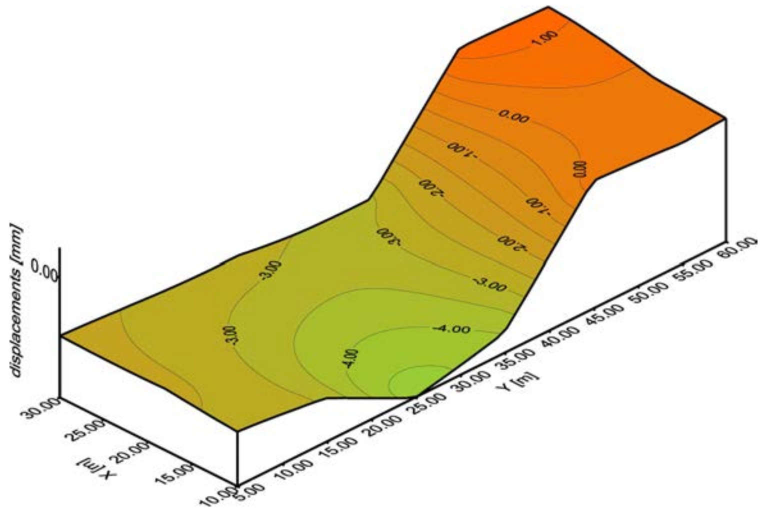

Fig. 5. Geometrical displacement model.

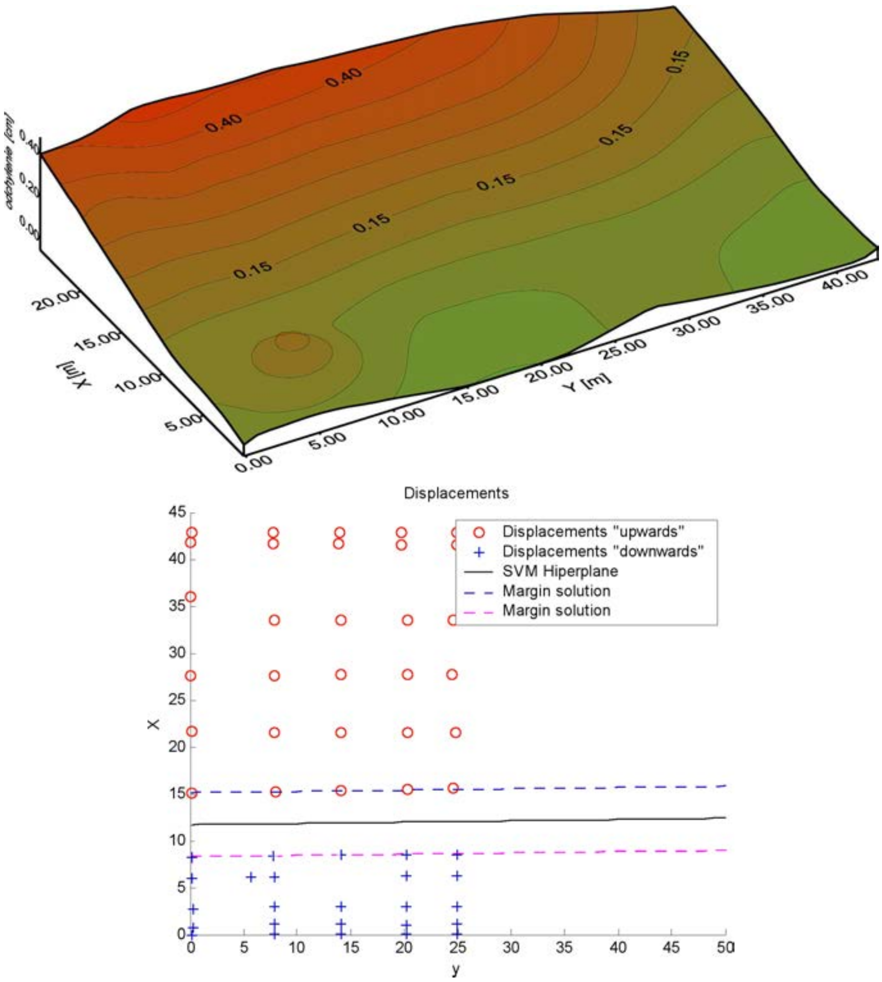

Fig. 6. Geometrical displacement model and classification of controlled points.

The comparison of the results obtained with the SVM method presented in Fig. 4 and those obtained with the classic method (Fig. 5), indicates an agreement between the identification of the points with "upwards" directed displacements and the identification of the points with "downwards" directed displacements.

The presented procedure of solving the classification task in the form of a two-value decision can be used as a starting point for the assessment of deviation of a surface in the horizontal and vertical plane in such objects as buildings and bridge and viaduct piers. The directions of the vectors of deviations determine the membership of characteristic points in either class 1 or class 2. The object for which displacements in the horizontal and vertical plane were assessed was Our Most Holy Lady Church in Toruń, whose walls deviat inward towards the nave as a result of the modernisation works on the roof. The classification of controlled points and a geometrical displacement model have been presented in Fig. 6.

6. Conclusions

The linear SVM network presented in the article can be used for classifying separable and inseparable data, and also for solving regression tasks. This network, similarly to other networks taught under monitoring (e.g. MLP networks), plays the role of a universal training data approximator. Its advantage is good generalisation and a relatively low sensitivity to the amount of training data, which is particularly important while solving tasks with a limited amount of data (e.g., in the case of measurements of displacements of civil engineering objects).

It is worth emphasising that the results obtained with SVM classification are in agreement with the results obtained with the traditional geodetic method used to determine vertical displacements. This approach makes it possible to determine an optimum hyperplane separating points in a measurement and control geodetic network into two classes, which can be a basis for making decisions about the type and place of possible work needed to protect an object subjected to uneven settlement and deviations in the horizontal and vertical plane.

References

[1] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[2] C. Burges. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining*. Ed. Sama Fayyad, Kluwer, 1–43, 2000.

[3] J. Gil. *Examples of applications of neural networks in geodesy* [in Polish]. Publishing House of the University of Zielona Góra, Zielona Góra, 2006.

[4] S.M. Gunn. *Support Vector Machines for Classification or Regression*. Technical Report, 1998.

[5] S. Haykin. *Neural networks, a comprehensive foundation*. Macmillan College Publishing Company, New York 1994.

[6] M. Mrówczyńska. A linear SVM network for determination of vertical displacement [in Polish: Sieć liniowa SVM do wyznaczenia przemieszczeń pionowych]. *Przegląd Geodezyjny* [Eng.: *Geodesic Review*], 3/2014, Warszawa, 2014.

[7] S. Osowski. *Neural networks for information processing* [in Polish]. Publishing House of Warsaw University of Technology, Warszawa, 2006.

[8] B. Schölkopf, A. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. Massachusetts Institute of Technology Press, Cambridge, 2001.

[9] I. Steinwart, A. Christmann. *Support Vector Machines*. Springer, 2008.

[10] V. Vapnik. *Statistical learning theory*. Wiley, New York, 1998.

[11] L. Zanni, T. Serafini, G. Zanghirati. Parallel Software for Training Large Scale Support Vector Machines on Multiprocessor Systems. *Journal of Machine Learning Research*, **7**: 1467–1492, 2006.