DE GRUYTER OPEN

Reports on Geodesy and Geoinformatics vol. 98 /2015; pages 18-27 DOI: 10.2478/rgg-2015-0002

THE SVM METHOD AS AN INSTRUMENT FOR THE CLASSIFICATION OF VERTICAL DISPLACEMENTS

Maria Mrówczyńska

Department of Land and Environment Engineering
Institute of Building Engineering, University of Zielona Góra
Poland

Abstract

The article presents the basic rules for constructing and training neural networks called the Support Vector Machine method as well as possible applications for this kind of network. SVM networks are mainly used for solving tasks of classifying linearly and non-linearly separable data and regression. However, in recent years more applications have been found for them. The networks also solve such problems as the recognition of signals and images as well as speech identification.
In this paper, non-linear SVM networks have been used for classifying linearly separable and non-separable data with a view to formulating a model of displacements of points in a measurement-control network. The points of the measurement-control network were placed on a civil engineering object located on expansive soil (linearly separable data) and represented a mining exploitation area (linearly non-separable data). The task of training SVM networks requires the use of quadratic programming in search of an optimum point of the Lagrangian function in relation to the parameters being optimised. In the case of linearly non-separable data, the SVM method makes it possible to find a hyperplane which classifies objects as correctly as possible, and at the same time is located possibly far away from concentrations typical of each class.

Keywords: neural networks, classification, vertical displacements.

1. Introduction

Measurements of displacements and deformations of engineering objects are intended to provide information about the behaviour of the whole object or its components. This information is obtained by transformations in a discrete set of displacement vectors determined for selected points of the object. This set is the basic product of geodetic displacement measurements, which makes it possible to determine displacement parameters for the object block, the approximation of the vector displacement area and its geometrical interpretation. Methods of calculating displacements consist of several stages. The first of them consists in checking the correctness of the observation material and its correction if necessary, then a reference base is found and a reference system is defined where displace components of controlled points are calculated. The final stage is the assessment of the significance of the displacements (Prószyński, Kwaśniak 2006). In practice a number of methods are used for determining displacements e.g. the method of observation differences (the forward differencing method), the method of coordinate differences, the method of combined adjustment.

While calculating the results of displacement and deformation measurements, mathematical models are used, usually quite complex ones. These models can be used to describe: an object under research, a phenomenon, a control network with a model of the observation or reference system error. Among the models only concerning the process of determining displacements, it is possible to mention static and kinematic network models. These models make it possible to determine displacements for specific time intervals as well as the location for particular points of time. SVM classifiers can be regarded as integrated models that make it possible to connect the analysis of the observations and the analysis of the results with the physical interpretation of the behaviour of the monitored object (Prószyński, Kwaśniak, 2006). This approach enables determination of a model of vertical displacements and a zone on the object where the senses of displacements change.

SVM (*Support Vector Machine*) classifiers, called the SVM technique, were invented by Vapnik (Vapnik, 1998) as a new approach to constructing and training networks. SVM networks belong to the group of unidirectional networks. They usually have a two-layer structure and can use different types of activation functions (Osowski, 2006; Zanni et al., 2006).

In comparison with MLP (*Multi-Layer Perceptron*) neural networks, SVM networks do not have drawbacks typical of MLP networks, i.e. the possibility that the minimisation process is stopped at one of the many local minima, or an arbitrarily adopted initial network architecture which determines future generalisation abilities of the network.

In general, in the case of linearly separable data, SVM networks make it possible to find a separation hyperplane separating two classes (Fig. 1). In the case of linearly non-separable data, the SVM method makes it possible to find a hyperplane which correctly classifies objects, and at the same time is located possibly far away from concentrations typical of each class (Mrówczyńska, 2015). In the case of linearly non-separable data, by raising dimensionality and with the use of the SVM method, it is possible to find a curvilinear separation boundary with a maximum separation margin. The paper discusses the problem of using an SVM network for classifying linearly non-separable data, which is exemplified by vertical displacements observed in the Legnica-Głogów Copper Mining Area. The vertical displacements were determined on the basis of the results of levelling measurements carried out in the years 1967 – 2008. The article also presents the classification of vertical displacements registered in a structure located on expansive soil.

2. A non-linear SVM network in a classification task

If we assume that a set of training pairs $(\mathbf{x}_i, d_i)$, $i = 1,\ldots,p$, is being classified, where $\mathbf{x}_i$ is the input vector and the assigned value d_i is 1 or -1, then we will write the equation of the hyperplane separating both classes as:

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0 = \sum_{j=1}^{N} w_j x_j + b = 0, \tag{1}$$

where the *N*-dimensional weight vector $\mathbf{w} = [w_1, w_2, \ldots, w_N]$, $\mathbf{x} = [x_1, x_2, \ldots, x_N]$, the weight b is polarization. The equation which makes it possible to decide which point is a member of a particular class has the form:

$$\begin{aligned} \mathbf{w}^T\mathbf{x}_i + b \geq 1 \rightarrow d_i = 1 \\ \mathbf{w}^T\mathbf{x}_i + b \leq 1 \rightarrow d_i = -1 \end{aligned}. \tag{2}$$

An optimum hyperplane $g(\mathbf{x}) = \mathbf{w}_o^T\mathbf{x} + b_o = 0$ is one for which there is a maximum separation margin (Fig. 1). The distance between the supporting vector $\mathbf{x}_{sv}$ and the hyperplane is expressed by the formula (Bishop 2006, Osowski 2006):

$$\tau = \frac{g(\mathbf{x}_{sv})}{\|\mathbf{x}\|} = \begin{cases} \dfrac{1}{\|\mathbf{w}\|} & \text{dla} \quad g(\mathbf{x}_{sv}) = 1 \\ -\dfrac{1}{\|\mathbf{w}\|} & \text{dla} \quad g(\mathbf{x}_{sv}) = -1 \end{cases}, \tag{3}$$

and the separation margin between two classes k is $k = 2\tau$.

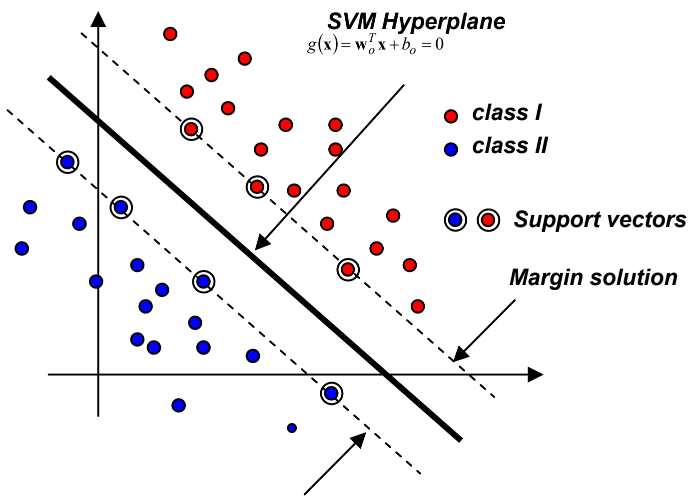

Fig. 1. An optimum hyperplane with a maximum separation margin

The support vectors are represented by the points that are the closest to the hyperplane. Because of their location, the support vectors are the most difficult to classify, but they determine the location of the hyperplane and the width of the separation margin.

In the case of linearly non-separable data, a common solution is to project the original data into a function space (a space of features), where the data become linearly separable with a probability close to 1. Usually, the dimension of the space of features *K* is much larger than the dimension of the space of the original N, the transformation of one space into the other is a non-linear transformation (Haykin, 1994; Cover, 1965). The linearly non-separable data in the two dimensional space of the original were transformed into the space of features which was defined by means of Gaussian functions. After the transformation, the data become linearly separable and can be divided with one separation hyperplane. The location of the separation hyperplane is determined in the space of features, and in the space of the original we only see its representation.

As a result of the transformation, the equation (1) is replaced with the equation of the hyperplane in the space of features, written as:

$$g(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b = 0 = \sum_{j=1}^{N} w_j \varphi_j(\mathbf{x}) + b = 0, \quad (4)$$

where w_j is a weight leading from a neuron in the hidden layer to the output neuron whose signal is defined by the equation:

$$y(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b. \quad (5)$$

Fig. 2 shows the architecture of the SVM network to which the output signal expressed by the dependence (5) corresponds. Analysing the SVM structure, it is possible to notice that it is a structure analogous to that of networks with radial base functions (RBF). The difference between a non-linear SVM network and a radial network is that the functions $\varphi(\mathbf{x})$ can assume any form (polynomial, radial sigmoidal).

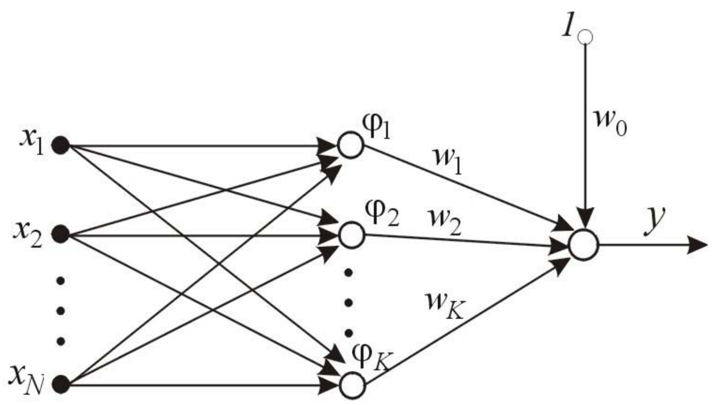

Fig. 2. The architecture of a non-linear SVM network

Training a non-linear SVM network consists in choosing its weights in such a way as to determine for linearly non-separable data an optimum hyperplane which minimises the possibility of a classification error. At the same time, the condition of maximisation of the separation margin has to be satisfied. While classifying linearly non-separable data, it is necessary to define a non-negative collative variable λ whose task is to decrease the current width of the separation margin. A problem presented in this way is called the primal problem reduced to the minimisation of a function:

$$\min\{\varphi(\mathbf{w},\boldsymbol{\lambda})\} = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{p}\lambda_i \tag{6}$$

with the linear constraints:

$$\begin{matrix} d_i\left(\mathbf{w}^T\varphi(\mathbf{x})_i + b\right) \geq 1 - \lambda_i \\ \lambda_i \geq 0 \end{matrix}, \tag{7}$$

where the parameter *C* is a certain constant, which is arbitrarily adopted by the user. It is worth emphasizing that the larger is the value of the parameter *C*, the narrower is the separation margin, and the smaller is the number of support vectors. In addition, when the value of the parameter *C* is small, the network almost works like a linear network. Hence the conclusion that the coefficient *C* works as a regulator between the width of the separation margin and the number of errors in the training set.

The primal (optimization) problem, when transformed to the dual problem, is a problem of square programming with linear constraints on weights, which is solved by means of Lagrangian multipliers through the minimisation of the Lagrangian function (Bishop, 2006; Gunn, 1998).

Solving the dual problem makes it possible to determine optimum Lagrangian multipliers which are the basis for determining optimum values of weights in the network according to the dependence:

$$\mathbf{w} = \sum_{i=1}^{p}\alpha_i d_i \varphi(\mathbf{x})_i \ . \tag{8}$$

The output signal of the non-linear SVM is finally expressed as:

$$y(\mathbf{x}) = \mathbf{w}^T\varphi(\mathbf{x}) + b = \sum_{i=1}^{P_{sv}}\alpha_i d_i K(\mathbf{x},\mathbf{x}_i) + b \ , \tag{9}$$

where P_{sv} is the number of support vectors $\mathbf{x}_i$, which is equal to the number of nonzero Lagrangian multipliers.

The function $K(\mathbf{x}_i,\mathbf{x}_j)$, in the formulation (9), is called a kernel function. This is a symmetrical function, which is the scalar product of the vector function $\varphi(\mathbf{x})$. The function $K(\mathbf{x}_i,\mathbf{x}_j)$ can be characterized by different forms of kernels: linear, polynomial, radial and sigmoidal (Mrówczyńska, 2015).

3. Numerical example

In order to use SVM classifiers, two sets of input data were made. They consisted of vertical displacements of controlled points situated on a building located on expansive soil (linearly separable data) and vertical displacements of points of a measurement-control network situated in the Legnica-Głogów Copper Mining Area (linearly non-separable data). The output data consist of information about the membership of points of the measurement-control network in one of the two classes.

In the case of the building located on expansive soil, it is necessary to take into account the fact that it is mainly exposed to vertical forces. In this case, the problem of geodesic measurements is focused on the research of foundation settlement. The object under research is represented by 11 controlled points stabilized on the walls of a building at the height of about 0.50 m from ground level. The layout of the points resulted from geological and engineering reasons. Each periodical measurement provided information about changes in the condition of the object in the form of 22 observations. During the observations the structure of the measurement-control network was the same. The acceptable error of measurement accuracy for relative displacements was not greater than ±0.3 mm.

Using the notion of own reference system, a displacement model for the controlled points was formulated, which was the basis for a bivalent decision about the linear separability of two classes of points with different senses of displacements (Mrówczyńska, 2014b). As a result of the use of the SVM method, three support vectors were determined, created at points No. 3,7,8, for which the Lagrangian coefficients were different from zero, and the classifier formed a separation margin of about 2 m (Fig. 3).

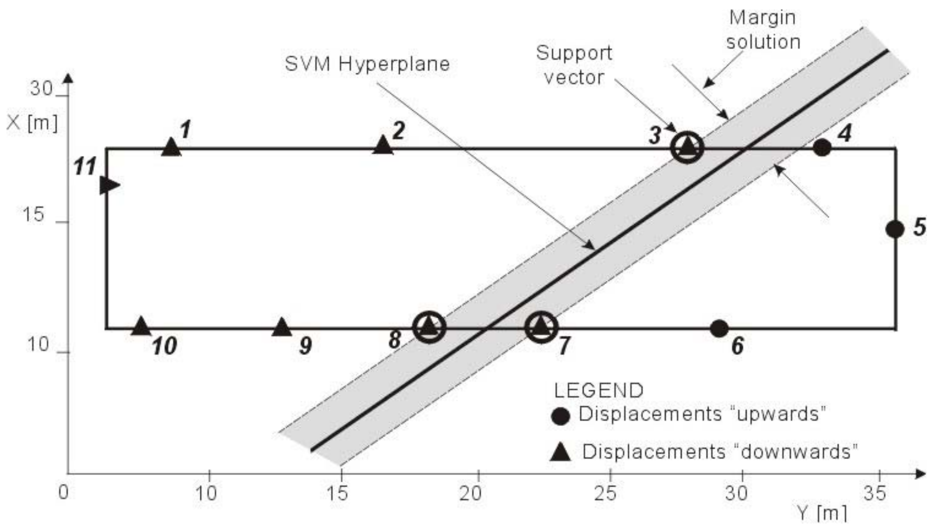

Fig. 3. The classification of points in the measurement-control network with the SVM method.

In order to compare the results obtained with the SVM method, a geometric displacement model of was created with the classic method, using an algorithm for the minimisation of absolute deviations and a criterion for a critical value of the increment of the square of the norm of the vector of corrections to the observations (Gil, 1995; Kuligowski, 1986)). A graphical interpretation of the displacement values that were determined is presented in Fig. 4.

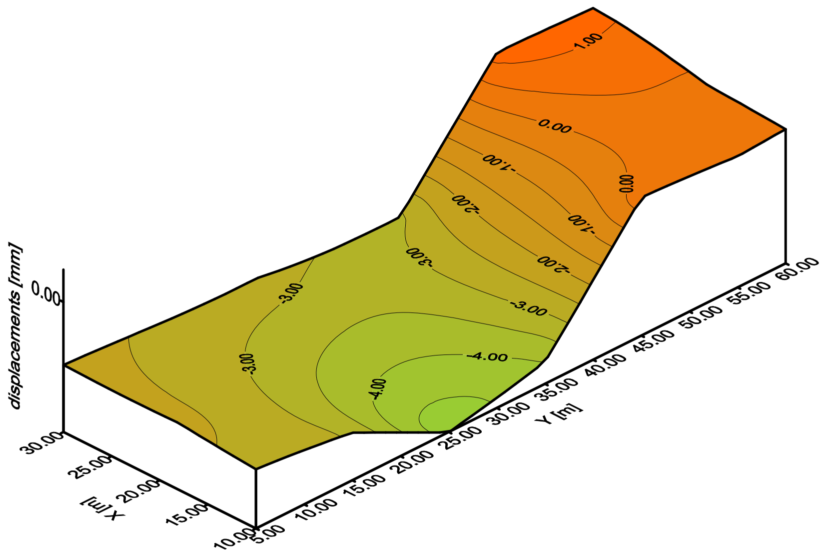

Fig. 4. A geometric displacement model

When the results obtained with the SVM method (Fig. 3) are compared with those obtained with the classic method (Fig. 4), it is possible to notice agreement between the identification of points whose displacements are directed "up" and those whose displacements are directed "down".

In the case of linearly non-separable data, the results of vertical displacement measurements carried out in the Legnica-Głogów Copper Mining Area were used. The Legnica-Głogów Copper Mining Area is located in the southern part of the Fore-Sudetic Monocline, where effects of mining exploitation can be felt. Detailed information on the geological structure of this area can be found in the work (Markiewicz, 2003).

The vertical displacements of the controlled points located in the Legnica-Głogów Copper Mining Area were determined through the analysis of the results of three measurement campaigns carried out in the years 1967 – 2008. In that area of 75000 ha, the measurement-control network adopted for the purpose of this paper had 218 points connected with one another with 302 observations (Fig. 5).

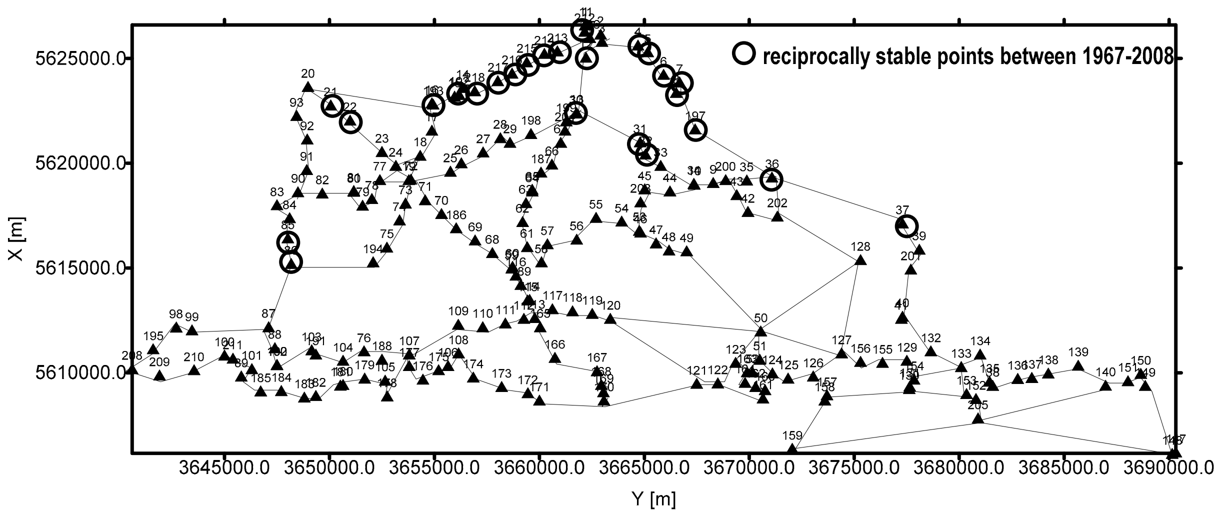

Fig. 5. A scheme of the measurement-control network

By means of the displacement model that was created (Mrówczyńska, 2014a), the membership of the controlled points in one of the two classes was determined, depending on the senses of the displacements. When information about the membership of the points in one of the two classes is available, the problem of the classifier boils down to finding a curvilinear separation boundary separating the two classes with a minimum error and a maximum separation margin. For this purpose a non-linear SVM network was used, where radial functions were used as kernel functions, and the parameter was assumed to be *C*=1000. As a result of the calculations, a curvilinear separation boundary based on 25 support vectors was determined (Fig. 6). The width of the separation margin ranges from 600 m to 1700 m.

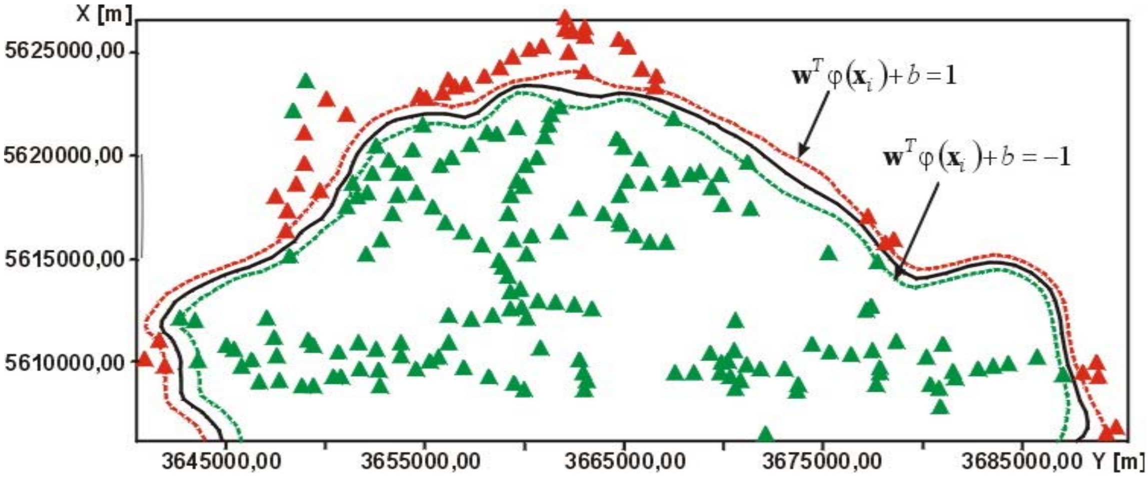

Fig. 6. A curvilinear separation boundary between points with different senses of displacements

The course of action presented in the paper is complemented with a geometric displacement model in the form of isolines, where the direction of the line of the greatest fall is marked, as well as the line of the curvilinear separation boundary (Fig. 7). The displacement values in Figure 7 are in millimetres.

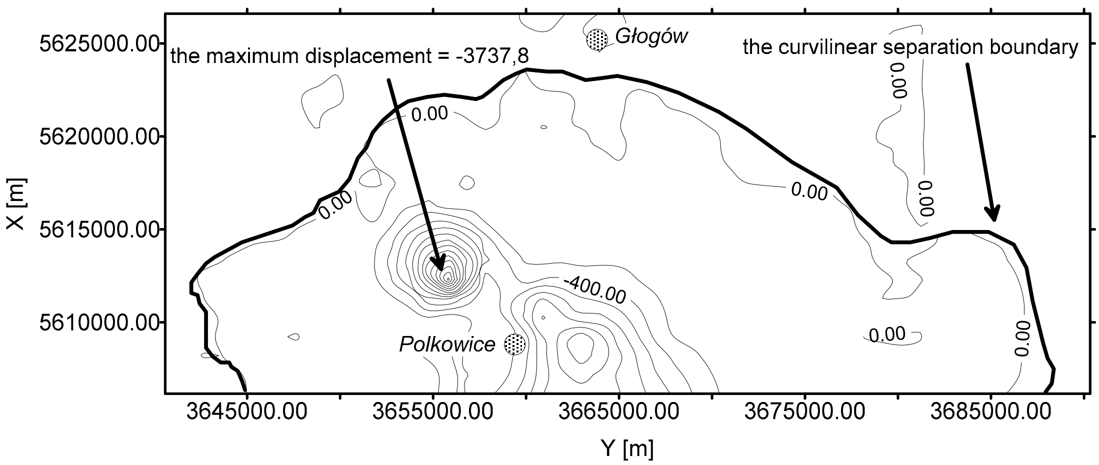

Fig. 7. A geometric displacement model obtained in the years 1967 – 2008

4. Conclusions

The non-linear SVM network presented in the paper was used for classifying linearly separable and non-separable data. These networks, like other networks trained under supervision (for example MLP networks), play the role of the universal approximator of training data. Their advantage is good generalization, which results from small sensitivity to the amount of training data. This is particularly important for solving problems in situations where the amount of data is limited (like in the case of measuring displacements and deformations).

The approach presented in the paper makes it possible to determine an optimum hyperplane in the space of features, which separates the points of the measurement control geodesic network into two classes. This can be the basis for making a decision about locating civil engineering objects in mining exploitation areas. The results that were obtained can also be helpful in making a decision about the place and method of possible protective work on civil engineering objects (e.g. determining the places where the foundations should be strengthened, the places and height at which clamps should be installed) where uneven settlement is detected. In the case of non-separable data the most important problem to solve is determination of the parameter C and the kind of the kernel function. In the classification of displacements in the Legnica-Głogów Copper Mining Area, the best results were obtained when a radial function was adopted as the kernel function and the parameter C=1000.

It is worth emphasizing that the results of the SVM classification agree with the results obtained with the traditional geodesic method of determining vertical displacements, and the line of the separation boundary is similar to the isolines of zero displacements.

References

Bishop C. M. (2006). Pattern Recognition and Machine Learningupport. *Springer*.

Cover T. (1965). Geometrical and statistical propertirs of system sof linear inequalities with applications in pattern recoqnition. *IEEE Trans. Electronic Computers, vol. 14.*

Gil J. (1995). Badanie nieliniowego geodezyjnego modelu przemieszczeń (na przykładzie obciążonego podłoża gruntowego). *Wydawnictwo Wyższej Szkoły Inżynierskiej w Zielonej Górze*. Zielona Góra.

Gunn S. M. (1998). Support Vector Machines for Classification or Regression. *Technical Report*.

Haykin S. (1994). Neural networks, a comprehensive foundation. *Macmillan College Publishing Company*. New York.

Kuligowski J. L. (1986). Zarys teorii grafów. *Wydawnictwo PWN*. Warszawa.

Markiewicz A. (2003). Halotektoniczne uwarunkowania sedymentacji i deformacji osadów kenozoicznych w południowej części Monokliny Przedsudeckiej (SW Polska*). Oficyna Wydawnicza Uniwersytetu Zielonogórskiego.* Zielona Góra.

Mrówczyńska M. (2014a). Klasyfikatory neuronowe typu SVM w zastosowaniu do klasyfikacji przemieszczeń pionowych na obszarze LGOM. *Zeszyty Naukowe SIGMiE PAN*. Kraków.

Mrówczyńska M. (2014b). Sieć liniowa SVM do wyznaczenia przemieszczeń pionowych. *Przegląd Geodezyjny 3/2014b*. Warszawa.

Mrówczyńska M. (2015). Studium nad doborem metod inteligencji numerycznej do rozwiązywania problemów z geodezji inżynieryjnej. *Oficyna Wydawnicza Uniwersytetu Zielonogórskiego*. Zielona Góra.

Osowski S. (2006). Sieci neuronowe do przetwarzania informacji. *Oficyna Wydawnicza Politechniki Warszawskiej.* Warszawa.

Prószyński W., Kwaśniak M. (2006). Podstawy geodezyjnego wyznaczenia przemieszczeń. Pojęcia i elementy metodyki. *Oficyna Wydawnicza Politechniki Warszawskiej*. Warszawa.

Vapnik V. (1998). Statistical learning theory. *Wiley*, New York.

Zanni L., Serafini T., Zanghirati. G. (2006). Parallel Software for Training Large Scale Support Vector Machines on Multiprocessor Systems. *Journal of Machine Learning Research 7, 1467-1492*.

Author:

Maria Mrówczyńska [1)], m.mrowczynska@ib.uz.zgora.pl
[1)] Department of Land and Environment Engineering
Institute of Building Engineering, University of Zielona Góra
Poland