Least Square Multi-classification Support Vector Machines: Pair-wise (P_ALS-MSVM) & Piece-wise (P_ILS-MSVM) formulations

OLUTAYO OLADUNNI and THEODORE B. TRAFALIS
School of Industrial Engineering
University of Oklahoma
202 West Boyd, Room 124 Norman, OK 73019
USA

Abstract: - This paper presents least square formulations for constructing a pair-wise linear and nonlinear classification decision functions. The formulations are based on the KKT system obtained from the optimality conditions of the P_AMSVM problem. This derivation can be considered a variant of the Suykens and Vandewalle's least square multi-class SVMs, with the notable difference been the equality constraints used in their problem formulation and the encoded classes (labels) represented by multiple streams for the system output. The least square formulation will be of two types namely the pairwise least square multi-classification support vector machine (P_ALS-MSVM) and piecewise least square multi-classification support vector machine (P_ILS-MSVM). A piece-wise MSVM formulation will be selected from the existing literature and its optimality conditions will be written out and expressed as a least square problem. The structures of both LS problems are essentially the same with the only difference been the matrix which has the dataset information. Since the Mercer conditions are applicable, kernels can be implemented as appropriate for nonlinear classification problems.

Key Words: - pairwise, piecewise, MSVM, multi-class, least square, linear system, classification

1 Introduction

Support Vector Machines (SVMs) developed by Vapnik [1] are based on statistical learning theory and have been successfully applied to a wide range of problems. It does not only represent a universal spirit of learning methodology, it complements the existing methodology of modeling and simulation of decision making processes for classification and regression problems. The prime advantage of the SVMs for classification problems is its ability to perform a mapping of the variables in a high (possibly infinite) feature space thus, providing an avenue for exploring kernel classifiers. Classification is done in this feature space by making use of a hyperplane of a nonlinear decision surface, i.e. optimal separating plane. In order to map the variables into a high feature space we make use of Mercer's condition.

Multi-classification SVMs is an extension of support vector machines (SVMs) involving three or more classes. There is active research in this area aiming at the construction of single optimization models, the reduction of the computational effort needed to solve the resulting large scale optimization problems and subproblems. Earlier attempts investigated solving *k* SVM models, where *k* is the number of classes and *k(k-1)/2* is the number of SVM classifiers [2, 3]. Other attempts involved the solution of a single optimization problem using all data at once [3, 4, 5]. Those attempts are arguably the most well-constructed multi-class formulations most closely aligned with Vapnik's principle [1] of always trying to solve problems directly.
Our main concern will be a multi-classification formulation which will be expressed as a single optimization problem.

We will look at the development of a pair-wise multi-classification support vector machine (P$_A$MSVM) expressed as a least square problem (P$_A$LS-MSVMs) and finally we will express a slight variant of the piece-wise multi-classification support vector machines formulation [3, 4, 5] (P$_I$LS-MSVM) as a least square piece-wise multi-classification support vector machine (P$_I$LS-MSVM). The P$_A$LS-MSVM is different from the LS-MSVM of Suykens and Vandewalle [6] in the sense that we use inequality constraints as in the primal P$_A$MSVMs formulation.

This paper is organized as follows. In section 2 pair-wise multi-classification support vector machine (P$_A$MSVM) models are discussed. In section 3 we present a least square pair-wise support multi-classification support vector machines formulation (LS-P$_A$MSVMs) and piece-wise multi-classification support vector machine formulation (LS-P$_I$MSVM) respectively. In section 4 we give computation results and application areas, and section 5 concludes the paper.

2 Pair-wise multi-classification support vector machines

In pairwise classification, we train a classifier for each possible pair of classes. For *k* classes, this results to *k(k-1)/2* SVM classifiers. Below is a pairwise MSVM formulation for a linearly separable problem:

$$\min_{w,\gamma} \left\{ \frac{1}{2}\sum_{i<j}^{k} \left\| w^{ij} \right\|^2 \mid y^{ij}(A^{ij}w^{ij} - e\gamma^{ij}) \geq e,\ i<j \right\} \tag{1}$$

Here is a 3 classes problem (*k* = 3) rewritten in matrix notation

Let

$$C = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}$$

$$A = \begin{pmatrix} A^1 & 0 & 0 \\ -A^2 & 0 & 0 \\ 0 & A^1 & 0 \\ 0 & -A^3 & 0 \\ 0 & 0 & A^2 \\ 0 & 0 & -A^3 \end{pmatrix} \quad E = \begin{pmatrix} -e^1 & 0 & 0 \\ e^2 & 0 & 0 \\ 0 & -e^1 & 0 \\ 0 & e^3 & 0 \\ 0 & 0 & -e^2 \\ 0 & 0 & e^3 \end{pmatrix}$$

Where $I \in R^{n \times n}$ is the identity matrix, $A^i \in R^{m_i \times n}$, $A^j \in R^{m_j \times n}$ $i<j$, and $e^i \in R^{m_i \times 1}$ $e^j \in R^{m_j \times 1}$ $i<j$ is a vector of ones.

So for *k* > 2 problem (1) can be expressed in the following form:

$$\begin{aligned} &\min_{w,\gamma} \ \frac{1}{2}\|Cw\|^2 \\ &s.t.\ Aw + E\gamma - e \geq 0 \end{aligned} \tag{2}$$

where

$$w = \left[w^{12^T}, w^{13^T}, .., w^{(k-1)k^T} \right]^T \ and\ \ \gamma = \left[\gamma^{12}, \gamma^{13}, .., \gamma^{(k-1)k} \right]^T$$

The constrained optimization is solved by introducing Lagrange multipliers $\alpha \geq 0$ and a Lagrangian

$$L(\alpha, w, \gamma) = \frac{1}{2}\|Cw\|^2 - \alpha^T(Aw + E\gamma - e) \tag{3}$$

Differentiating the Lagrangian with respect to w *and* γ leads to

$$\frac{dL}{dw} = (C^TC)w - A^T\alpha = 0 \tag{4}$$

$$\frac{dL}{d\gamma} = -E^T\alpha = 0 \tag{5}$$

To eliminate the variables $w\ and\ \gamma$ from the Lagrangian, matrix (C^TC) needs to be an invertible matrix. Since (C^TC) is invertible, we have

$$(C^TC)^{-1}A^T = A^T \tag{6}$$

So from equations (4) & (5) and (6) we can obtain the relations

$$\begin{aligned} &w = (C^TC)^{-1}A^T\alpha = A^T\alpha \\ &0 = E^T\alpha \end{aligned} \tag{7}$$

3 Least Square Multi-classification Support Vector Machines

3.1 LS-P$_A$MSVM Formulation

The derivation of the P$_A$LS-MSVM formulation is based on a slight variation of problem (2), where we minimize the regularization term and the sum of square error referring to slack variables that account for inseparability of the constraints of (2).

$$\begin{aligned}&\min_{w,\gamma}\ \frac{1}{2}\|Cw\|^2+\lambda\frac{1}{2}\xi^T I\xi\\ &s.t.\ \ Aw+E\gamma-e+I\xi\geq 0\\ &\qquad \xi\geq 0\end{aligned}\qquad (8)$$

Define the Lagrangian

$$L(\alpha,\beta,w,\gamma,\xi)=\left\{\begin{matrix}\frac{1}{2}\|Cw\|^2+\lambda\frac{1}{2}\xi^T I\xi\\ -\alpha^T(Aw+E\gamma-e+I\xi)\\ -\beta^T(\xi)\\ \end{matrix}\right\}\qquad (9)$$

where $\alpha\geq 0$ are Lagrange multipliers.

From the optimality conditions, we obtain the Karush-Kuhn Tucker (KKT) system by differentiating the Lagrangian with respect to $\alpha,\beta,\xi,w\ and\ \gamma$

$$\frac{dL}{dw}=(C^TC)w-A^T\alpha=0\qquad (10)$$

$$\frac{dL}{d\gamma}=-E^T\alpha=0\qquad (11)$$

$$\frac{dL}{d\xi}=\lambda I\xi-\alpha-\beta=0\qquad (12)$$

$$\frac{dL}{d\alpha}=-(Aw+E\gamma-e+I\xi)=0\qquad (13)$$

$$\frac{dL}{d\beta}=-\xi=0\qquad (14)$$

Eliminating variables $w\ and\ \xi$ and substituting the relations obtained

$$w=(C^TC)^{-1}A^T\alpha=A^T\alpha\qquad (15)$$

$$\xi=\lambda^{-1}I^{-1}(\alpha+\beta)\qquad (16)$$

into the optimality conditions to obtain the linear system

$$AA^T\alpha+E\gamma-e+I\{\lambda^{-1}I^{-1}(\alpha+\beta)\}=0\qquad (17)$$

$$\lambda^{-1}I^{-1}(\alpha+\beta)=0\qquad (18)$$

$$\begin{pmatrix}0 & E^T & 0\\ E & AA^T+\lambda^{-1}I & \lambda^{-1}I\\ 0 & \lambda^{-1}I & \lambda^{-1}I\end{pmatrix}\begin{pmatrix}\gamma\\ \alpha\\ \beta\end{pmatrix}=\begin{pmatrix}0\\ e\\ 0\end{pmatrix}\Rightarrow\ A_{ls}x_{ls}=b_{ls}\qquad (19)$$

A_{ls} is a symmetric matrix, $A_{ls}\in R^{m\times m}$, the $A_{ls}x_{ls}=b_{ls}$ system is non-homogeneous and consistent.

3.2 LS-P$_I$MSVM Formulation

This is the MSVM formulation of Bredensteiner and Bennet [4] which declares a set of points $A^i, i=1,\cdots,k$ represented by matrices $A^i\in R^{m_i\times n}$ are piecewise-linearly separable if there exist $w^i\in R^n$ and $\gamma^i\in R$ such that:

$$A^iw^i-\gamma^i e>A^iw^j-\gamma^j e,\ \ i,j=1,\ldots,k,\ \ i\neq j\qquad (20)$$

In canonical form

$$x^T(w^i-w^j)-e(\gamma^i-\gamma^j)>1,\ \ i,j=1,\ldots,k,\ \ i\neq j\qquad (21)$$

The bounding plane separating classes *i* and *j* is defined $x^T(w^i-w^j)=e(\gamma^i-\gamma^j)$.

Below is a piecewise MSVM formulation for a linearly separable problem:

$$\begin{aligned}&\min_{w,\gamma}\ \frac{1}{2}\sum_{i<j}^{k}\left\|w^i-w^j\right\|_2^2+\frac{1}{2}\sum_{i=1}^{k}\left\|w^i\right\|_2^2\\ &s.t.\ \ A^i(w^i-w^j)-e(\gamma^i-\gamma^j)-e\geq 0,\\ &\qquad i,j=1,\cdots,k\quad i\neq j\end{aligned}\qquad (22)$$

To classify a new point x we employ the “Max Wins” strategy [2, 3].

if sign [$x^T(w^i-w^j)-e(\gamma^i-\gamma^j)$] says that x is in the *i*th class,

then the vote for the *i*th class is increased by one. Otherwise, the *j*th is increased by one. Then we predict x as being in the class with the largest vote. In the case that those two classes have identical votes, select the one with the smallest index. Alternatively, we could compute $g_i(x)=x^Tw^i-\gamma^i$, and find i such that $g_i(x)=x^Tw^i-\gamma^i$ is maximized i.e. $g(x)=\max_{i=1,..,k}g_i(x)$, where $g(x)$ is a decision function.

From relation (20) and (21) the class with the maximum vote should also be the class with $\max_{i=1,..,k}g_i(x)$.

The derivation of the least square P$_I$MSVM is based on a slight variation of the Bredensteiner and Bennet [4] MSVM formulation, where the norm is minimized simultaneously with the sum of square error given in problem (23).

$$\begin{aligned}&\min_{w,\gamma}\ \frac{1}{2}\sum_{i<j}^{k}\left\|w^i-w^j\right\|^2+\frac{1}{2}\sum_{i=1}^{k}\left\|w^i\right\|^2+\lambda\frac{1}{2}\sum_{\substack{i=1\\ j\neq i}}^{k}\sum_{j=1}^{k}(\xi^{ij})^2\\ &s.t.\ \ A^i(w^i-w^j)-e(\gamma^i-\gamma^j)-e+\xi^{ij}\geq 0,\\ &\qquad \xi^{ij}\geq 0\\ &\qquad i,j=1,\cdots,k\quad i\neq j\end{aligned}\qquad (23)$$

Here is a 3 classes problem (*k* = 3) rewritten in matrix notation

Let

$$\hat{C} = \begin{pmatrix} I & -I & 0 \\ I & 0 & -I \\ 0 & I & -I \end{pmatrix}$$

$$\hat{A} = \begin{pmatrix} A^1 & -A^1 & 0 \\ A^1 & 0 & -A^1 \\ -A^2 & A^2 & 0 \\ 0 & A^2 & -A^2 \\ -A^3 & 0 & A^3 \\ 0 & -A^3 & A^3 \end{pmatrix} \quad \hat{E} = \begin{pmatrix} -e^1 & e^1 & 0 \\ -e^1 & 0 & e^1 \\ e^2 & -e^2 & 0 \\ 0 & -e^2 & e^2 \\ e^3 & 0 & -e^3 \\ 0 & e^3 & -e^3 \end{pmatrix}$$

where identity matrix $I \in R^{n\times n}$, $A^i \in R^{m_i \times n}$ and $e^i \in R^{m_i \times 1}$ is a vector of ones. So when $k > 2$, we simply adjust the matrices $\hat{C}$, $\hat{A}$ and $\hat{E}$, and express problem (23) in the matrix notation

$$\min_{w,\gamma} \ \frac{1}{2}\left\|\hat{C}w\right\|^2 + \frac{1}{2}\left\|w\right\|^2 + \lambda\frac{1}{2}\xi^T I\xi$$
$$s.t. \ \hat{A}w + \hat{E}\gamma - e + I\xi \geq 0 \qquad (24)$$
$$\xi \geq 0$$

By defining the Lagrangian for problem (24)

$$L(\alpha,\beta,w,\gamma,\xi) = \begin{Bmatrix} \frac{1}{2}\left\|\hat{C}w\right\|^2 + \frac{1}{2}\left\|w\right\|^2 + \lambda\frac{1}{2}\xi^T I\xi \\ -\alpha^T(\hat{A}w + \hat{E}\gamma - e + I\xi) - \beta^T(\xi) \end{Bmatrix} \qquad (25)$$

where $\alpha \geq 0$ are Lagrange multipliers, we can obtain the Karush-Kuhn Tucker (KKT) system by differentiating the Lagrangian with respect to $\alpha, \beta, \xi, w \ and \ \gamma$

$$\frac{dL}{dw} = (I + \hat{C}^T\hat{C})w - \hat{A}^T\alpha = 0 \qquad (26)$$

$$\frac{dL}{d\gamma} = -\hat{E}^T\alpha = 0 \qquad (27)$$

$$\frac{dL}{d\xi} = \lambda I\xi - \alpha - \beta = 0 \qquad (28)$$

$$\frac{dL}{d\alpha} = -(\hat{A}w + \hat{E}\gamma - e + I\xi) = 0 \qquad (29)$$

$$\frac{dL}{d\beta} = -\xi = 0 \qquad (30)$$

Eliminating variables $w \ and \ \xi$ in eqs. (25) - (30) we obtain

$$w = (I + \hat{C}^T\hat{C})^{-1}\hat{A}^T\alpha = \frac{1}{k+1}\hat{A}^T\alpha \qquad (31)$$

$$\xi = \lambda^{-1}I^{-1}(\alpha + \beta) \qquad (32)$$

and substituting into the optimality conditions we obtain the linear system below:

$$\frac{1}{k+1}\hat{A}\hat{A}^T\alpha + \hat{E}\gamma - e + I\{\lambda^{-1}I^{-1}(\alpha+\beta)\} = 0 \qquad (33)$$

$$\lambda^{-1}I^{-1}(\alpha + \beta) = 0 \qquad (34)$$

In matrix form

$$\begin{pmatrix} 0 & \hat{E}^T & 0 \\ \hat{E} & \frac{1}{k+1}\hat{A}\hat{A}^T + \lambda^{-1}I & \lambda^{-1}I \\ 0 & \lambda^{-1}I & \lambda^{-1}I \end{pmatrix}\begin{pmatrix} \gamma \\ \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ e \\ 0 \end{pmatrix} \Rightarrow \hat{A}_{ls}\hat{x}_{ls} = \hat{b}_{ls} \qquad (35)$$

Where $\hat{A}_{ls}$ is a symmetric matrix, $\hat{A}_{ls} \in R^{m\times m}$, the $\hat{A}_{ls}\hat{x}_{ls} = \hat{b}_{ls}$ system is non-homogeneous and consistent.

To solve the linear system (19) and (35), we express the system as a least square unconstrained minimization problem (unified approach) [8] of the form

$$f(x) = \frac{\alpha_{ls}}{2}x^T x + \frac{1}{2}\left\|(b - A\,x\,)\right\|^2 \qquad (36)$$

where α_{ls} is the trade off constant that seeks to find a compromise between obtaining the minimum norm solution and minimum residual solution. This form is especially useful for rank deficient/singular matrices and badly conditioned Hessians. In case that the Hessian is not positive definite (P.D.) we force the matrix or Hessian to become positive definite for a suitable $\alpha_{ls} > 0$, and determine a solution x that minimizes $f(x)$.

$$f(x) = \frac{\alpha_{ls}}{2}x^T x + \frac{1}{2}\left(b^T b - 2b^T Ax + x^T A^T Ax\right) \qquad (37)$$

The optimality conditions for $f(x)$ are as follows:

First order conditions

$$\nabla f(x) = \alpha_{ls}Ix - A^T b + A^T Ax = 0 \qquad (38)$$

Second order conditions

$$\nabla^2 f(x) = A^T Ax + \alpha_{ls}Ix\,, \qquad (39)$$

where $\nabla^2 f(x)$ is P.D.

The minimum solution is obtained from the first order relation

$$A^T Ax + \alpha_{ls}Ix = A^T b \ \Rightarrow \ (A^T A + \alpha_{ls}I)x = A^T b \qquad (40)$$
$$x = (A^T A + \alpha_{ls}I)^{-1}A^T b$$

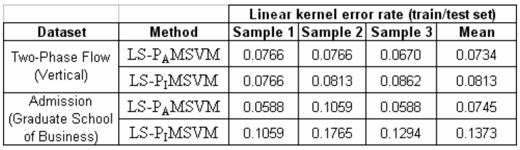
4 Numerical Testing

In this section we present the computational results using systems (19) and (35) for discriminating between *k* classes. Experiments were carried out on a vertical 1 inch two-phase flow dataset [9] and the admission data for graduate school of business [10]. Description of datasets is as follows:

Vertical Two-Phase Flow Dataset: The two-phase flow dataset [9] uses a pair of flow rates (superficial gas and liquid velocity) to delineate the flow regime. There are 209 instances (points) and 2 attributes (features). The distribution of instances with respect to their class is as follows: 44 instances in class 1 (bubble flow), 102 instances in class 2 (intermittent flow), and 63 instances in class 3 (annular flow).

Admission Data for Graduate School of Business: The admission data dataset [10] uses the undergraduate grade point average (GPA) and graduate management aptitude test (GMAT) scores to help determine which applicants should be admitted to the school's graduate program. There are 85 instances (points) and 2 attributes (features). The distribution of instances with respect to their class is as follows: 28 instances in class 1 (not admitted), 26 instances in class 2 (borderline), and 31 instances in class 3 (admitted).

Both formulations were implemented using the optimization and matrix decomposition routines in the MATLAB [11] software. The two-phase flow dataset was scaled by taking the natural logarithm of each instance, while the admission data were scaled using the Z – score normalization. The methods were trained on 50% of the dataset, and tested on the whole dataset (50% training, 50% testing data), all randomly drawn from the dataset to obtain 3 training sample data. We report the results of the LS formulation of P_ALS-MSVM (λ, $\alpha_{ls} = 1$), and the least square formulation of P_ILS-MSVM (λ, $\alpha_{ls} = 1$). The most accurate model will be the one that has the highest classification accuracy.



		Linear kernel error rate (train/test set)			
Dataset	Method	Sample 1	Sample 2	Sample 3	Mean
Two-Phase Flow (Vertical)	LS-P_AMSVM	0.0766	0.0766	0.0670	0.0734
	LS-P_IMSVM	0.0766	0.0813	0.0862	0.0813
Admission (Graduate School of Business)	LS-P_AMSVM	0.0588	0.1059	0.0588	0.0745
	LS-P_IMSVM	0.1059	0.1765	0.1294	0.1373

Table 1: Performance of LS-P_AMSVM and LS-P_IMSVM.

Table 1 contains the results for three multi-class methods, P_ALS-MSVM, and P_ILS-MSVM on the Two-Phase flow and Admission dataset. The misclassification error of the P_ILS-MSVM is noticeably higher than the P_ALS-MSVM. The mean error rate of the P_ALS-MSVM on the Two-Phase flow dataset is lower and produces the best generalization ability. The mean accuracy of the P_ILS-MSVM is also acceptable (see Table 1). On the Admission dataset, the mean accuracy of the P_ALS-MSVM is lower and produces the best generalization ability. The results demonstrate the potential of the methods. Note that the quality of the solution is dependent on the choice of λ and α_{ls}. For this problem λ, $\alpha_{ls} = 1$ were sufficient enough to present good results. Further computational studies could be of interest such as varying choices of λ and α_{ls} to determine the effect on the performance of the methods.

5 Conclusion and Future Work

In this paper we presented an extension of the binary SVM to the multiclass SVM. We have derived a least square formulation using the KKT system obtained from the optimality conditions for both the pairwise and piecewise least square SVM. The two methods proposed present an accurate and capable alternative to existing earlier methods based on solving *k* SVM models which can be tiresome due to the number of SVM models one would have to solve in order to discriminate between *k* classes.

The proposed methods were applied to the Two-Phase flow and Admission dataset and comparisons were made between the two methods, P_ALS-MSVM and P_ILS-MSVM. The results are very encouraging considering that the kernel used is a linear kernel that indicates that the both datasets are close to be linearly separable with a tolerable misclassification rate. All error rates are acceptable, with the analysis of both the Two-Phase flow and the analysis of the Admission data favoring the P_ALS-MSVM as being the best most accurate method. The P_ALS-MSVM has shown consistency with respect to the dataset of interest. Further studies will involve the use of nonlinear kernel classifiers. Since, the Mercer's condition is applicable kernel functions can be incorporated into the MSVM formulations.

Further computational studies would include the implementation of several preprocessing schemes to normalize the data. Also varying the choices of λ and α_{ls} to help determine the effect on the performance of the model.

Reference:

[1] Vapnik, V. *Statistical Learning Theory.* John Wiley & Sons, Inc., **1998**.

[2] Santosa, B.; Conway, T.; Trafalis, T. B. Knowledge Based-Clustering and Application of Multi-Class SVM for Genes Expression Analysis, *Intelligent Engineering Systems through Artificial Neural Networks* **2002**, 12, 391 – 395.

[3] Hsu, C-W.; Lin, C-J. A Comparison of Methods for Multi-class Support Vector Machines, *IEEE Transactions on Neural Networks* **2002** 13, 415 – 425.

[4] Bredensteiner E. J.; Bennet, K. P. Multicategory Classification by Support Vector Machines, *Computational Optimization and Applications* **1999**, 12, 53 – 79.

[5] Weston, J.; Watkins, C. Multi-class Support Vector Machines, In M. Verleysen, editor, *Proceedings of ESANN99*, Brussels, D. Facto Press, **1999**.

[6] Suykens, J. A. K.; Vandewalle, J. Multiclass Least Squares Support Vector Machine Classifers, In *Proc. of the Int. Joint Conf. on Neural Networks (IJCNN'99)* **1999c**, Washington, DC.

[7] Gestel, T. V.; Suykens, J. A. K.; Baesens B.; Viane, S.; Vanthiennen, J.; Dedene, G.; De Moor, B.; Vandewalle, J. Benchmarking Least Squares Support Vector Machine Classifers. *Machine Learning* **2004**, 54, 5 – 32.

[8] Lewis, J. M.; Lakshmivarahan, S.; Sudarshan, D. *Dynamic Data Assimilation*; Cambridge University Press (manuscript to be published).

[9] Trafalis, T. B.; Oladunni, O.; Papavassiliou, D. V. Two-Phase Flow Regime Identification with a Multi-Classification SVM Model, School of Industrial Engineering, College of Engineering, University of Oklahoma. (accepted) to appear in the *Industrial & Engineering Chemistry Research*.

[10] Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistics Analysis*; Prentice Hall: New Jersey, **2002**.

[11] *MATLAB User's Guide*. The Math-Works, Inc., Natwick, MA 01760, 1994-2003. http://www.mathworks.com.

[12] Chang, C-C.; Lin, C-J. *LIBSVM: A Library for Support Vector Machines* **2001**, http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[13] Hsu, C-W.; Chang, C-C.; Lin, C-J. *A Practical Guide to Support Vector Classification*, Technical Report Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan **2003.**