The use of LS-SVM in the classification of brain tumors based on Magnetic Resonance Spectroscopy signals

L. Lukas[1], **A. Devos**[1], **J.A.K. Suykens**[1], **L. Vanhamme**[1], **S. Van Huffel**[1], **A.R. Tate**[2], **C. Majós**[3], **C. Arús**[4]

[1] K.U. Leuven, Dept. of Electrical Engineering ESAT-SISTA, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
[2] St. George's Hospital Medical School, Dept. of Biochemistry, London, United Kingdom
[3] Institut de Diagnòstic per la Imatge (IDI) - Centre Bellvitge Autovia de Castelldefels, km. 2.7, L'Hospitalet de Llobregat, 08907, Spain
[4] Universitat Autonòma de Barcelona, Departament de Bioquímica i Biologia Molecular, Cerdanyola del Vallès, Spain
{llukas, andy.devos, johan.suykens, leentje.vanhamme, sabine.vanhuffel}@esat.kuleuven.ac.be

Abstract. Least Squares Support Vector Machines (LS-SVM) have been developed and successfully applied to classification problems in many areas. In comparison with several other classical methods this technique consistently performs very well on a large variety of problems. Here, results on the application of LS-SVM for classification of brain tumors based on Magnetic Resonance Spectroscopy (MRS) signals are presented. Several kernels are used and compared to find the optimal classifier. Despite the high dimensionality and the scarcity of the input data, and the fact that no additional clinical information is used, a good ROC and classification performance can be achieved after applying leave-one-out cross-validation for hyperparameter selection together with an additional bias term correction. The improvement of this classification based on MRS signals will lead to an advanced tool for the discrimination of brain tumors, which is presently under development for the INTERPRET project [3].

Keywords. Classification, Magnetic Resonance Spectroscopy, Brain tumors, Least Squares Support Vector Machines.

1 Introduction

Since the introduction of Support Vector Machines (SVM) [11], there has been a growing interest in kernel-based methods and many successes have been reported. In this paper we focus on the use of Least Squares SVM classifiers

(LS-SVM) [7]. In [10] it has been shown that LS-SVM classifiers consistently perform very well on a large variety of problems in comparison with a large amount of different methods. LS-SVMs lead to linear systems instead of QP problems and sparseness and robustness might be imposed where needed. In comparison with classical neural networks (LS)-SVM perform well in high dimensional feature spaces. This motivates the use of LS-SVM classifiers on the classification of brain tumors based on Magnetic Resonance (MR) Spectroscopy (S) signals [6, 8]. Results are reported on a database which is still under development. First results indicate that LS-SVM classifiers with RBF kernel perform well, even on raw data without prior expert knowledge.

This paper is organized as follows. In Section 2 we briefly explain the concepts of Magnetic Resonance Spectroscopy. In Section 3 we give a brief introduction on LS-SVM. Section 4 presents the application of LS-SVM to the classification of two types of brain tumors using MRS signals.

2 Magnetic Resonance Spectroscopy

(Nuclear) Magnetic Resonance (NMR) is a non-invasive technique that has been used to acquire spatially resolved images of living organisms and to monitor changes in the metabolism [1]. The most frequently used application of MR in humans and animals is MR Imaging (MRI) in which anatomical information is obtained. Another application of clinical MR is MRS in which chemical information can be obtained from a well-defined region in for example the human brain. Various localization techniques have been developed to measure MRS signals. Single-voxel techniques acquire information from one spatially resolved location, called a voxel. The result of a single-voxel MRS measurement is a time-domain signal, which is visualized as a spectrum after a Fourier transformation. In this spectrum peaks corresponding to different chemical substances can be observed. In proton MR spectra of normal and pathological brain resonances from e.g. N-acetyl containing compounds (NAc), choline, creatine, alanine, lactate, lipids and residual water can be observed. Recent studies indicate that proton MR spectra of brain tumors show differences in spectral patterns that correlate with the tumor type and grade [6, 8]. This motivates the research described in this paper.

3 Least Squares Support Vector Machines

Given a training set of N data points $\{x_k, y_k\}_{k=1}^N$, where $x_k \in \mathbb{R}^n$ is the k-th input pattern and $y_k \in \{-1, +1\}$ the corresponding given class label, the LS-SVM classifier [7] takes the form

$$y(x) = \text{sign}[\sum_{k=1}^{N} \alpha_k \, y_k \, K(x, x_k) + b] \tag{1}$$

where α_k are called support values and b is a constant and are the solution to a linear system (in contrast with standard SVMs which are related to quadratic

programming). $K(\cdot,\cdot)$ is a positive definite kernel, often chosen as $K(x,x_k) = x_k^T x$ (linear SVM); $K(x,x_k) = (x_k^T x + c)^d$ (polynomial SVM of degree d); $K(x,x_k) = \exp\{-\|x-x_k\|_2^2/\sigma^2\}$ (RBF kernel).

The classifier is obtained as follows. A binary classification problem is translated into the following set of inequalities $w^T\varphi(x_k) + b \geq +1$ if $y_k = +1$ and $w^T\varphi(x_k) + b \leq -1$ if $y_k = -1$, which can be written as

$$y_k[w^T\varphi(x_k) + b] \geq 1, \qquad k = 1, ..., N. \tag{2}$$

Here $\varphi(\cdot)$ is a nonlinear mapping of the input space into a higher dimensional space which can be infinite dimensional. In the primal weight space the classifier takes the form $y(x) = \text{sign}[w^T\varphi(x) + b]$ but is never evaluated in this way. Instead (1) is applied, where α, b are obtained as follows.

One formulates the optimization problem

$$\min_{w,b,e} \mathcal{J}_{LS}(w,b,e) = \frac{1}{2}w^T w + \gamma\,\frac{1}{2}\sum_{k=1}^{N} e_k^2 \tag{3}$$

subject to the equality constraints

$$y_k\,[w^T\varphi(x_k) + b] = 1 - e_k,\; k = 1, ..., N. \tag{4}$$

The Lagrangian is $\mathcal{L}(w,b,e;\alpha) = \mathcal{J}_{LS} - \sum_{k=1}^{N}\alpha_k\{y_k[w^T\varphi(x_k)+b] - 1 + e_k\}$ with Lagrange multipliers $\alpha_k \in \mathbb{R}$ (called support values). The conditions for optimality are given by $\frac{\partial\mathcal{L}}{\partial w} = 0$, $\frac{\partial\mathcal{L}}{\partial b} = 0$, $\frac{\partial\mathcal{L}}{\partial e_k} = 0$, $\frac{\partial\mathcal{L}}{\partial\alpha_k} = 0$ which gives $w = \sum_{k=1}^{N}\alpha_k y_k\varphi(x_k)$, $\sum_{k=1}^{N}\alpha_k y_k = 0$, $\alpha_k = \gamma e_k$, $y_k[w^T\varphi(x_k)+b]-1+e_k = 0$, for $k = 1, ..., N$. Elimination of w and e gives

$$\left[\begin{array}{c|c} 0 & Y^T \\ \hline Y & \Omega + \gamma^{-1}I \end{array}\right]\left[\begin{array}{c} b \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline 1_v \end{array}\right]. \tag{5}$$

with $Y = [y_1; ...; y_N]$, $1_v = [1; ...; 1]$, $e = [e_1; ...; e_N]$, $\alpha = [\alpha_1; ...; \alpha_N]$. Mercer's theorem is applied to the matrix Ω with

$$\begin{aligned} \Omega_{kl} &= y_k y_l\,\varphi(x_k)^T\varphi(x_l) \\ &= y_k y_l\,K(x_k,x_l). \end{aligned} \tag{6}$$

Hence one chooses a positive definite kernel $K(\cdot,\cdot)$ that satisfies Mercer condition.

4 LS-SVM classification of brain tumors

The dataset considered in this paper consists of 47 MR spectra covering two classes, corresponding to two different brain tumor types, i.e. 23 spectra of glioblastomas and 24 spectra of meningiomas, labeled as class 1 and class 2, respectively. The raw data are acquired in the time domain at the centre IDI (Institut de Diagnòstic per la Imatge), contractor of the INTERPRET project [3]. The signals were acquired at 1.5 T (Philips) using the PRESS pulse

sequence (TR/TE=2000/136 ms). The results shown in this paper are still preliminary due to the fact that the database is still under development, but on the other hand it already gives a good indication of what might be achievable. A few preprocessing steps are carried out: alignment and phase correction with Klose's method [4] and filtering of the dominating residual water peak using HLSVD [5]. The signals are normalized by dividing by the intensity of the water peak (using a spectrum without water suppression) [9]. The resulting signal is transformed to the frequency domain by a FFT transformation. For each signal the magnitude spectrum is considered only in the frequency region of interest (-34 Hz to -240 Hz), corresponding to 107 input variables. A few resonance peaks of chemical substances that may be indicative for brain tumors are the peaks of choline (-93 Hz), creatine (-105 Hz), NAc (-167 Hz), the doublet of alanine (central frequency of -203 Hz), the doublet of lactate (central frequency of -215 Hz) and a lipid peak (-218 Hz).

The mean magnitude spectra on Figure 1 show the peak pattern characteristics of the two brain tumor classes. Some typical features which distinguish the glioblastomas from the meningiomas are (1) the presence of lipid peaks in glioblastomas, (2) the presence of lactate doublets in glioblastomas, (3) the presence of alanine peaks in meningiomas. The difficulty of distinguishing the two classes given this dataset is due to the limited number of validated spectra available.

LS-SVM classifiers are designed and three kernels, namely linear, polynomial and RBF kernels have been tried. The LS-SVM input vector consists of 107 variables which correspond to the frequency-domain magnitude values of the frequency region of interest. For each kernel, the hyperparameters are tuned to achieve the best leave-one-out (LOO) generalization performance, taking into account the small size of the data set. The LS-SVM with RBF kernel gives $\gamma = 4, \sigma = 1$ after LOO selection with 91.5 percent correct classification. The polynomial kernel with $c = 2$ and degree $d = 2$ gives 78.7 percent correct classification. With $\gamma = 0.078$, the LS-SVM with linear kernel also gives 78.7 percent correct classification. These performances are all obtained by applying first LOO selection to find the optimal tuning parameters and then applying a bias term correction $y(x) = \text{sign}[\sum_{k=1}^{N} \alpha_k \, y_k \, K(x, x_k) + b + \Delta b]$ where Δb is chosen such that optimal classification accuracy is achieved. This Δb correction has been done on the training set due to the fact of the small data set size. However, the risk for overfitting is minimal because of the LOO tuning parameter selection and given the fact that LS-SVM classifiers with RBF kernel incorporate principles of both parametric and non-parametric regularization.

To judge the discrimination ability of the LS-SVM classifier for the brain tumor dataset, Receiver Operating Characteristic (ROC) curves [2] have been constructed. ROC curves are often used to evaluate the discrimination between benign and malignant cases. In Figure 2, the ROC curves are shown. The area under the ROC curves for the three kernels are 0.9438 (RBF kernel), 0.7862 (linear kernel), and 0.8170 (polynomial kernel). Hence the LS-SVM with RBF kernel outperforms the choice of the other kernels. An interesting finding from the mentioned experiment is that, although the given dataset is scarce and has high dimensionality, the LS-SVM classifier achieves more than 90 percent of correct classification after LOO hyperparameter selection, even

without incorporating any prior knowledge of the clinical information.

5 Conclusions

LS-SVM have been applied to classify brain tumors based on MR signals. The application shows that LS-SVM classifiers are able to learn quite well from the raw input MR spectra in a high dimensional space, even without any additional clinical information about the patient or the acquisition of the MR signal. RBF kernels perform significantly better than linear kernels on this case. Further research will be based on augmenting the database, incorporating additional expert knowledge and relevant feature selection methods.

Acknowledgements

This research work was carried out at the ESAT laboratory and the Interdisciplinary Center of Neural Networks ICNN of the Katholieke Universiteit Leuven, in the framework of the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture (IUAP P4-02 & IUAP P4-24), the Concerted Action Project MEFISTO of the Flemish Community, the FWO project G.0407.02 and the IDO/99/03 project. AD research financed by IWT grant of the Flemish Institute for the promotion of scientific-technological research in the industry. LVH and JS are a postdoctoral researcher with the National Fund for Scientific Research FWO - Flanders. The EU (IST) programme has provided financial support for this work (INTERPRET IST-1999-10310).

References

[1] Gadian D.G., *NMR and its applications to living systems, 2nd ed.*, Oxford University Press, 1995.

[2] Hanley A.J., McNeil B.J., "The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve," *Radiology*, 143(1):29-36, 1982.

[3] International network for Pattern Recognition of Tumours Using Magnetic Resonance http://carbon.uab.es/INTERPRET/interpret_home.shtml

[4] Klose U., "In vivo proton spectroscopy in presence of eddy currents," *Magn. Reson. Med.*, 14:26-30, 1990.

[5] Pijnappel W.W.F., van den Boogaert A., de Beer R., van Ormondt D., "SVD-based quantification of magnetic resonance signals," *J. Magn. Reson.*, 97:122-134, 1992.

[6] Preul M.C., Caramanos Z., Collins D.L., Villemure J.-G., Leblanc R., Olivier A., Pokrupa R., Arnold D.L. "Accurate, non-invasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy," *Nature Medicine*, 2:323-325, 1996.

[7] Suykens J.A.K., Vandewalle J., "Least Squares support vector machine classifiers," *Neural Processing Letters*, 9(3):293-300, 1999.

[8] Tate A.R. *et al.*, "Towards a method for automated classification of ^{1}H MRS spectra from brain tumours," *NMR Biomed*, 11:177-191, 1998.

[9] Usenius J.P., Tuohimetsä S., Vainio P., Ala-Korpela M., Hiltunen Y., Kauppinen R.A., "Automated classification of human brain tumors by neural network analysis using in vivo 1H magnetic resonance spectroscopic metabolite phenotypes," *Neuroreport* 7: 1597-1600, 1996.

[10] Van Gestel T., Suykens J., Baesens B., Viaene S., Vanthienen J., Dedene G., De Moor B., Vandewalle J., "Benchmarking Least Squares Support Vector Machine Classifiers," ESAT-SISTA Technical Report 00-37 ESAT-SISTA, K.U.Leuven Belgium, 2000 (downloadable from www.esat.kuleuven.ac.be/sista/members/suykens.html).

[11] Vapnik V., *The nature of statistical learning theory*, Springer-Verlag, 1995

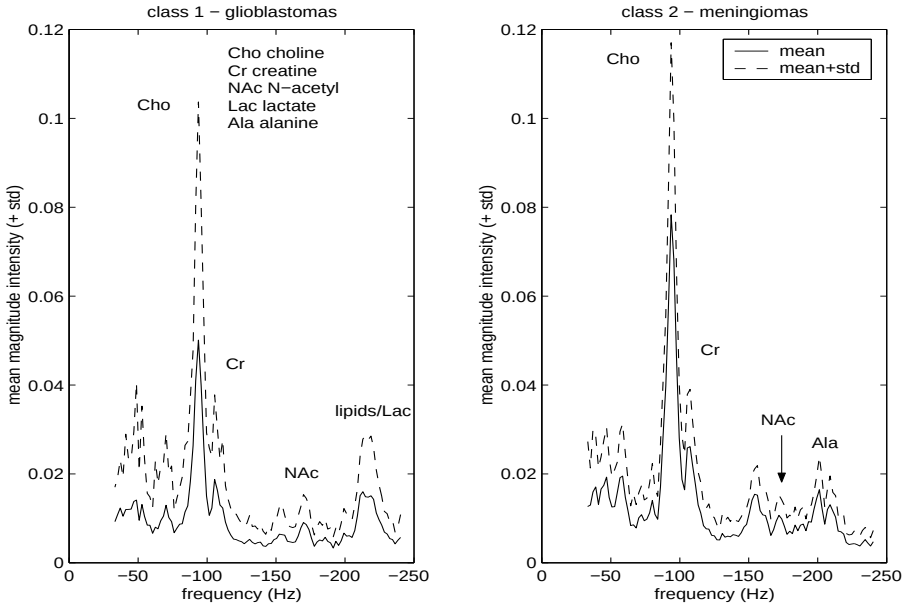

Figure 1: *Mean magnitude frequency spectra of the considered two classes: class 1 (left), class 2 (right) correspond to the glioblastomas and meningiomas, respectively. The solid lines are the means, while the dotted lines are the means plus the standard deviations of each class.*

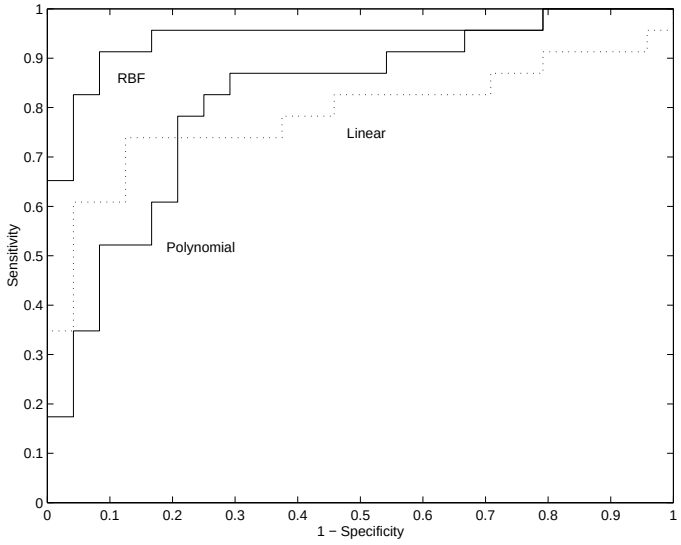

Figure 2: *ROC curves of LS-SVM classifiers for the given brain tumor classification problem. The Area Under the Curve (AUC) is 0.9438 for the RBF kernel, 0.7862 for the linear kernel, and 0.8170 for the polynomial kernel.*