A model for a complex polynomial SVM kernel

Dana Simian

Abstract

The aim of this paper is to present many computational aspects related to polynomial spaces of w-degree. We obtained algorithms for computing the dimension of homogeneous spaces of w-degree n and the exponents of the monomial basis of these spaces. These elements are very important for using these spaces as interpolation spaces. We discuss aspects regarding the complexity of these algorithms and implementation details.

1 Introduction

Support Vector Machines (SVMs) represent a class of neural networks, introduced by Vapnik ([3]). We will make a brief presentation of SVMs for binary classification in section 2. We limit our discussion to the binary case. There are many methods to make generalizations for several classes. In the recent years, SVMs have become a very popular tool for machine learning tasks and have been successfully applied in classification, regression and novelty detection. Many applications have been done in various fields: particle identification, face identification, text categorization, bioinformatics, database marketing.

Generally, the task of classification is to find a rule, which based on external observations assign an object to one of several classes. A classification task supposes the existence of training and testing data given in the form of data instances. Each instance in the training set contains one target value, named class label and several attributes named features. The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are given only the attributes. SVM solutions are characterized by a convex optimization problem. If the data is separable we obtain an optimal separating hyperplane with a maximal margin ([1]). In order to avoid the difficulties for the non separable data the idea of kernel substitution is used. The kernel methods, transform data into a feature space F, that usually has a huge dimension ([3]). With a suitable choice of kernel the data can become separable in feature space despite being non-separable by a hyperplane in the original input space. The basic properties of a kernel function are derived from Mercer's theorem ([1]). Under certain conditions, kernel functions can be interpreted as representing the inner product of data objects implicitly mapped into a nonlinear feature space. The "kernel trick" is to calculate the inner product in the feature space without knowing explicit the mapping function.

2 Support Vector Machines and kernels

An SVM algorithm can solve the problem of binary or multiclass classification. Many real-life data sets involve multiclass classification. We consider in this section only the problem of binary classification be cause there are many known methods to generalize the binary classifier to an n - class classifier ([1], [3]).

Let be given the data points $x_i \in R^d$, $i \in \{1, \ldots, m\}$ and their labels $y_i \in \{-1, 1\}$. We look for a function f which associates to each input data x its correct label $y = f(x)$. This function is named decision function and represents a hyperplane which divides the input data into two regions:

$$f(x) = sign(\langle w, x \rangle + b), \tag{1}$$

where $w = \sum_{i=1}^{m} \alpha_i x_i$. If the data set is separable then the conditions for classification without training error are $y_i(\langle w, x_i \rangle + b) > 0$. To maximize the margin the task is

$$min \leq \left(\frac{1}{2} \parallel w \parallel^2 \right),$$

subject to the constraints

$$y_i(\langle w, x_i \rangle + b) \geq 0, \ \forall i \in \{1, \ldots, n\}$$

The Wolfe dual problem requires the maximization with respect to α_i, of the function

$$W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \tag{2}$$

subject to the constraints

$$\alpha_i \geq 0, \quad \sum_{i=1}^{m} \alpha_i y_i = 0, \tag{3}$$

with α_i Lagrange multipliers (hence $\alpha_i \geq 0$).

3   Main results. Our model

We want to build and analyze a multiple kernel starting from the simple polynomial kernel and having 2 parameters, the degree d and the coefficient r.

3.1   Representation of the multiple kernel

Our chromosome is composed from 78 genes: 2 genes for each operation, 2 genes for the kernel's type, 4 genes for the degree parameter d_i, 12 genes for r_j. If the associated kernel is not polynomial, the last 16 genes are used to represent the real value of parameter γ_i, in place of d_i and r_i.

References

[1] C. Campbell , *An Introduction to Kernel Methods Radial Basis Function Network: Design and Applications*, Springer Verlag, Berlin, 2000.

[2] L. Diosan, M. Oltean, Rogozan A., J. P. Pecuchet, Improving svm performance using a linear combination of kernels, *Adaptive and Natural Computing Algorithms, ICANNGA07*, volume 4432 of LNCS, 2007, 218 - 227.

[3] Vapnik V., *The Nature of Statistical Learning Theory*, Springer, 1995.

Dana Simian
University Lucian Blaga of Sibiu
Faculty of Sciences
5-7 I. Ratiu str., Sibiu, 550021
ROMANIA
E-mail: *dana.simian@ulbsibiu.ro*