

One-Versus-One and One-Versus-All Multiclass SVM-RFE for Gene Selection in Cancer Classification

Kai-Bo Duan¹, Jagath C. Rajapakse^{1,2}, and Minh N. Nguyen¹

¹ BioInformatics Research Centre, School of Computer Engineering
Nanyang Technological University, Singapore

² Singapore-MIT Alliance, Singapore
{askbduan, asjagath, nmnguyen}@ntu.edu.sg

Abstract. We propose a feature selection method for multiclass classification. The proposed method selects features in backward elimination and computes feature ranking scores at each step from analysis of weight vectors of multiple two-class linear Support Vector Machine classifiers from one-versus-one or one-versus-all decomposition of a multi-class classification problem. We evaluated the proposed method on three gene expression datasets for multiclass cancer classification. For comparison, one filtering feature selection method was included in the numerical study. The study demonstrates the effectiveness of the proposed method in selecting a compact set of genes to ensure a good classification accuracy.

1 Introduction

Large number of genes together with relatively small number of samples is one characteristic of gene expression data available for cancer classification. This characteristic makes feature selection (FS) a necessary procedure for cancer classification with gene expression data to ensure reliable and meaningful classification results along with other benefits such less data storage and computation cost.

One category of multiclass classification methods are based on the combination of two-class (binary) classifiers. One-versus-one (OVO) or one-versus-all (OVA) binary classifiers are constructed to separate one class from another or to separate one class from all other classes respectively. These binary classifiers are then combined to conduct multiclass classification prediction. OVO and OVA are also two ways to extend popular Support Vector Machine (SVM) [1][2] classification method for multiclass classification. For multiclass classification methods that are based on OVO or OVA binary classifiers, one simple feature selection strategy probably is to select feature variables for each of the binary classifiers separately by using existing feature selection methods for two-class classification. In this way, different binary classifiers actually are constructed in different feature sub-spaces. However, combination of these binary classifiers is infeasible if the combination strategy is based on the distance of a sample to decision

boundaries of different binary classifiers. Unfortunately, the decision function of SVMs (before taking the sign function) corresponds to such a distance measurement and commonly used *Winner-Takes-All* combination strategy [3] for one-versus-all binary SVMs is primarily based on that distance measurement.

Specially for OVO and OVA multiclass SVMs, we will develop in this paper a feature selection method which is based on OVO or OVA binary SVM classifiers and assume all binary classifiers in combination take the same set of features as input variables. The proposed methods can be viewed as a multiclass generalization of SVM-RFE [4] feature selection method for two-class classification. SVM-RFE selects feature in backward elimination and uses coefficients of the weight vector of a binary linear SVM to compute feature ranking scores. The proposed feature selection method selects features using a backward elimination procedure similar to that of SVM-RFE but computes feature ranking scores at each step from analysis of the weight vectors of all OVO or OVA binary linear SVMs from OVO or OVA decomposition of a multiclass classification problem. The method will be evaluated on three multiclass gene expression datasets for cancer classification and will also be compared with a filtering feature selection method.

The rest of the paper is organized as follows. Section 2 gives a short description of two-class SVMs and, OVO and OVA multiclass SVMs. Section 3 describes the SVM-RFE feature selection for two-class classification; Section 4 presents the proposed multiclass SVM-RFE feature selection method; Section 5 is about the numerical study on three multiclass gene expression datasets for cancer classification. Section 6 contains the conclusions and some discussions.

2 SVMs

2.1 Two-Class SVMs

Support Vector Machines (SVMs) [1] [2] are one of the most popular supervised classification methods due to its superior classification performance in many applications. SVMs are also fairly insensitive to the *curse of dimensionality* and efficient in handling classification problems of large scale in both samples and input variables. SVMs was originally developed for two-class classification. First, input sample vectors \mathbf{x} are mapped from input space to a so-called *feature space* via a mapping function $\Phi(\cdot) : \mathbf{z} = \Phi(\mathbf{x})$, where \mathbf{z} denotes a vector in the feature space. Then, in the feature space, an optimal hyperplane discrimination function $f(\mathbf{z}) = \mathbf{w} \cdot \mathbf{z} + b$ is constructed to separate samples from two classes. The mapping function is implicitly defined by a kernel function which computes the inner-product of vectors in the feature space: $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{z}_i \cdot \mathbf{z}_j = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$. Commonly used kernel functions are linear, polynomial and Gaussian [1]. SVMs with a linear kernel function are referred to as linear SVMs.

2.2 Multiclass SVMs

Extension of SVMs from two-class to multiclass basically falls into two categories. The first category multiclass SVMs consider all samples and all classes