A regression perspective of binary and multi-class support vector machines

Patrick Groenen

Past President of International Association for Statistical Computing (IASC)

Erasmus University Rotterdam

Gertjan van den Burg

Alan Turing Institute

Abstract

Support vector machines (SVMs) have become a standard tool for binary classification problems that has become increasingly popular. In the machine learning literature, the SVM is often explained through the dual of a convex optimization problem. Instead, we approach it as a regression problem with a specific error function and a ridge type penalty term.

However, in the case that more than two classes need to be predicted often a series of binary SVMs are performed (one-versus-all or between all pairs of classes, one-versus-one). A disadvantage of such methods is that they are heuristics that do not simultaneously estimate all parameters in a single model. We discuss a new multiclass SVM loss function (GenSVM) that is based on a geometric representation of each class by a vertex of a simplex in $K - 1$ dimensional space. As with the binary SVM, an object that is predicted to be nearest to its class receives a zero error and if the object is closer to another class the error consists of a function of the distance to the zero-error region. The present approach is flexible in the hinge function that is used for calculating the error. It builds on the Huberized hinge errors that have as special cases the linear and quadratic hinges. It is also flexible in how these errors are added: we propose to use the L_p norm of the Huberized hinge error. This general loss function has the binary SVM and some existing multiclass SVM loss functions as special cases.

We discuss a majorization algorithm (also named MM or CCCP) that minimizes GenSVM. We present some numerical comparisons showing that for medium sized problems GenSVM compares with the best approaches.