Preprint #04-4

VISUALIZATION FOR CLASSIFICATION PROBLEMS, WITH EXAMPLES USING SUPPORT VECTOR MACHINES

by

Dianne Cook, Doina Caragea, Vasant Honavar
Iowa State University

February 10, 2004

ABSTRACT

In the simplest form support vector machines (SVM) define a separating hyperplane between classes generated from a subset of cases, called support vectors. The support vectors "mark" the boundary between two classes. The result is an interpretable classifier, where the importance of the variables to the classification, is identified by the coefficients of the variables defining the hyperplane. This paper describes visual methods that can be used with classifiers to understand cluster structure in data. The study leads to suggestions for adaptations to the SVM algorithm and ways for other classifiers to borrow from the SVM approach to improve the result. The end result for any data problem is a classification rule that is easy to understand with similar accuracy.