Feature Import Vector Machine: A General Classifier with Flexible Feature Selection

Samiran Ghosh\textsuperscript{1,2,*}, Yazhen Wang\textsuperscript{3}

\textsuperscript{1} Department of Family Medicine & Public Health Sciences
\textsuperscript{2} Center of Molecular Medicine and Genetics, Wayne State University
\textsuperscript{3} Department of Statistics, University of Wisconsin, Madison

\*Contact author: sghos@med.wayne.edu

\textbf{Keywords:} Classification, Import Vector Machine, Radial Basis Function, Regularization, Support Vector Machine

The support vector machine (SVM) and other reproducing kernel Hilbert space (RKHS) based classifier systems are drawing much attention recently due to its robustness and generalization capability. General theme here is to construct classifiers based on the training data in a high dimensional space by using all available dimensions. The SVM achieves huge data compression by selecting only few observations which lie close to the boundary of the classifier function. However when the number of observations are not very large (small $n$) but the number of dimensions/features are large (large $p$), then it is not necessary that all available features are of equal importance in the classification context. Possible selection of a useful fraction of the available features may result in huge data compression. In this paper we propose an algorithmic approach by means of which such an optimal set of features could be selected. In short, we reverse the traditional sequential observation selection strategy of SVM to that of sequential feature selection. To achieve this we have modified the solution proposed by Zhu and Hastie (2005) in the context of import vector machine (IVM), to select an optimal sub-dimensional model to build the final classifier with sufficient accuracy.