Using SVM to Estimate and Predict Large Dimensional Binary Choice Models

Yoosoon Chang (Indiana University), Guo Yan (University of Melbourne) ,and Joon Y. Park (Indiana University)

Abstract: The support vector machine (SVM) behaves like the quasi-maximum likelihood estimator (QMLE) asymptotically, although strictly it cannot be viewed as a QMLE, for the binary outcome observed with its covariates generated by the binary choice model (BCM). The QMLE is known to yield a consistent estimator for the slope parameter of the BCM, if the conditional expectation of covariates on its systematic component is given as a linear function. This can be readily established if we assume that a pseudo-true parameter value exists. Under the same condition, we show without assuming the existence of a pseudo-true value that the slope coefficient of the hyperplane provided by the SVM is a consistent estimator for the slope parameter of the BCM as long as two classes defined by a binary outcome are not severely imbalanced. In most cases, the SVM is thus asymptotically equivalent to the logistic regression. The finite sample performances of the two estimators could be quite distinct depending upon the distributions of covariates and errors, but no one dominates the other. The intercept parameter of the BCM may be consistently estimated, once a consistent estimator for its slope parameter is obtained.