

Evaluation of Optimal Linear Discriminant Function by 100-fold cross validation

Shuichi Shinmura

Faculty of Economics, Seikei Univ., Japan Shinmura@econ.seikei.ac.jp

Abstract

I develop new linear discriminant function called '**Revised IP-OLDF**' based on MNM criterion. It is compared with Fisher's linear discriminant function (LDF), quadratic discriminant function (QDF), logistic regression and soft margin SVM (S-SVM) by 100 fold cross validation. One hundred re-sampling data sets are generated from four kinds of original data such as: Fisher's Iris data (15 models), Swiss bank note data (16 models), CPD (Cephalo Pelvis Disproportion) data (19 models) and student data (31 models). The mean of error rates of 81 models of these methods are computed by LINGO and JMP. **Revised IP-OLDF** can recognize linear separable (MNM=0) data correctly, and can avoid the cases on $f(\mathbf{x}_i)=0$. MNM is the lower limit of NM for the training bootstrap samples. Two models of (Logi-MNM) and three models of (SVM-MNM) are negative. This result shows logistic regression and S-SVM can't avoid the cases on $f(\mathbf{x}_i)=0$. The mean error rates of Revised IP-OLDF are better than LDF, logistic regression and S-SVM.

Keywords: Linear Discriminant Function, Logistic Regression, Minimum Number of Misclassifications, Quadratic Discriminant Function, Revised IP-OLDF, Soft margin SVM
It is concluded that Revised IP-OLDF is better than other methods. In addition to these results, LDF and QDF never recognize linear separable data.