Chapter 6

Best Model for Swiss Banknote Data

- Explanation 1 of Matroska Feature-Selection Method (Method 2) –

Shuichi Shinmura
Shuichi Shinmura: Seikei Univ. Faculty of Economics, Tokyo, Japan, shinmura@econ.seikei.ac.jp

Abstract. When we discriminate Swiss banknote data by an optimal linear discriminant function (optimal LDF) using integer programming (IP-OLDF), we find this data is a linearly separable data (LSD). Because we examine all possible combination models, we find two variables model such as (X4, X6) is the minimum linearly separable model. Sixteen models including these two variables are linearly separable by the monotonic decrease of MNM (MNMk ≥ MNMk+1), and other forty-seven models are not linearly separable. Therefore, we compared eight LDFs by the best models with the minimum mean of error rate in the validation sample (M2) and obtain good results. Although we could not explain the useful meaning of the 95% confidence interval (CI) of discriminant coefficient until now, the pass/fail determinations using exam scores give us clear understanding by normalizing the coefficient. Seven LDFs become trivial LDFs. Only Fisher’s LDF is not a trivial LDF. Seven LDFs are Revised IP-OLDF based on a minimum number of misclassification (minimum NM, MNM), Revised LP-OLDF, Revised IPLP-OLDF, three SVMs and logistic regression. We are successful explaining the meaning of coefficient. Therefore, we discuss the relation between the best model and coefficient more precisely by Swiss banknote data.

We had studied the discrimination of LSD by Swiss banknote data, Japanese automobile data and six pass/fail determinations using exam scores precisely. When we discriminate six microarray data those are LSD, only Revised IP-OLDF can make feature selection naturally and reduce the high-dimensional gene space to the little gene space drastically. In the gene analysis, we call all linearly separable models as Matroska. The full model is the biggest Matroska that includes all smaller Matroskas in it. Because we had already known, the smallest Matroska can explain the structure of Matroska completely by the monotonic decrease of MNM. We propose the Matroska feature selection method for the microarray data (Method 2). Because the discrimination of LSD is not popular now, we explain the Method 2 by the detail examples of Swiss banknote data and Japanese automobile data in Chapter 7. On the other hand, the LASSO tries to make feature selection. If it cannot find the smallest Matroska in the data, it cannot explain the structure of Matroska. Swiss banknote data, Japanese automobile data, and six microarray data sets are useful evaluating the usefulness of other feature selection methods including LASSO.

Keywords: 100-fold Cross-validation for small Samples (Method 1), 95% CI of coefficient, Best Model, Matroska feature selection method (Method 2), Fisher’s Linear Discriminant Function (Fisher’s LDF), Logistic Regression, Support Vector Machines (SVMs), Revised IP-OLDF, Revised LP-OLDF, Revised IPLP-OLDF, LASSO.

6.1 Introduction

6.2 Swiss Banknote Data

6.2.1 Data Outlook

6.2.2 Comparison of Seven LDF for the Original Data

6.3 100-Fold Cross-Validation for Small Sample Method

6.3.1 Best Model Comparison

6.3.2 95% CI of Discriminant Coefficient

6.3.2.1 Consideration of 27 Models
6.3.2.2 Revised IP-OLDF
6.3.2.3 Hard-Margin SVM (H-SVM) and Other LDFs
6.4 Explanation 1 for Swiss Bank Note Data

6.4.1 Matryoshka in Linearly Separable Data

6.4.2 Explanation 1 of Method 2 by Swiss Banknote Data

6.5 Summary

REFERENCES
