
Discrete Discriminant Analysis: the performance of combining models

Anabela Marques¹, Ana Sousa Ferreira², and Margarida Cardoso³

¹ ESTBarreiro, Setúbal Polytechnic, Portugal, CEAUL, Project of Multivariate Data Analysis and Modelling, anabela.marques@estbarreiro.ips.pt

² LEAD, FPCE, University of Lisbon, Portugal, CEAUL, Project of Multivariate Data Analysis and Modelling, asferreira@fpce.ul.pt

³ Dep. of Quantitative Methods, ISCTE Business School, Lisbon, Portugal, UNIDE, margarida.cardoso@iscte.pt

Summary. The idea of combining models in Discrete Discriminant Analysis (DDA) is present in a growing number of papers which aim to obtain more robust and more stable models than any of the competing ones. This seems to be a promising approach since it is known that different DDA models perform differently on different subjects [1]. This is a particularly relevant issue when the groups are not well separated, which often occurs in practice. Recently, a new methodological approach was proposed based on a linear combination of the First-order Independence Model (FOIM) and the Dependence Trees Model (DTM) ([3] and [2]). In the present work we further explore the referred approach. Since FOIM assumes that the P discrete predictive variables are independent in each group and DTM takes the predictors relationships into account, we think that the proposed approach could be successfully applied to many real situations. In order to evaluate its performance, we consider both real and simulated data. Furthermore we present comparisons with alternative models performance. According to the training sample size the leave-one-out approach, v -fold cross validation or assessing the error rate in a test sample are considered. The MATLAB software is used for the algorithms' implementation.

Key words: Combining models, Dependence Trees model, Discrete Discriminant analysis, First Order Independence model.

References

1. I. Brito, G. Celeux, and A. Sousa Ferreira. Combining methods in supervised classification: A comparative study on discrete and continuous problems. *REV-STAT - Statistical Journal*, 4(3):201-225, 2006.
2. G. Celeux, and J. P. Nakache. Analyse Discriminante sur Variables Qualitatives. G. Celeux et J. P. Nakache Éditeurs, *POLYTECHNICA*, 1994.
3. A. Marques, A. Sousa Ferreira, and M. Cardoso. Combining Models in Discrete Discriminant Analysis in the Multiclass Case. In FEUP, editors, *Book of Abstract - COMPSTAT*, pages 270. FEUP, 2008.