

Local Classification of Discrete Variables by Latent Class Models

Michael Bucker¹

1. Fakultät Statistik, Technische Universität Dortmund, D-44221 Dortmund, Germany, buecker@statistik.tu-dortmund.de

Keywords: Local classification, Latent Class Analysis, Categorical data, Finite mixtures, Naive Bayes

Global classifiers may fail to distinguish classes adequately in discrimination problems with inhomogeneous groups. Instead, local methods that consider latent subclasses can be adopted in this case. Three different models for local discrimination of categorical variables are implemented in the `lcda` (latent class discriminant analysis) package. They are based on Latent Class Models (cf. [2]), which are discrete finite mixture distributions. Therefore, they can be estimated via the EM algorithm. One model is constructed analogously to the Mixture Discriminant Analysis (cf. [1]) by class conditional Latent Class Models. Two other techniques are based on the idea of Common Components Models (cf. [3]). Applicable model selection criteria and measures for the classification capability are suggested. In a simulation study, discriminative performance of the methods is compared to that of decision trees and the Naive Bayes classifier. It turns out that the MDA-type classifier can be seen as a localization of the Naive Bayes method. Additionally the procedures have been applied to a SNP (single nucleotide polymorphism) data set.

References

- [1] T. Hastie and R. Tibshirani. Discriminant analysis by gaussian mixtures. *Journal of the Royal Statistical Society B*, 58:155-176, 1996.
- [2] P.F. Lazarsfeld and N.W. Henry. *Latent structure analysis*. Houghton Mifflin, Boston, 1968.
- [3] M.K. Titsias and A.C. Likas. Shared kernel models for class conditional density estimation. *IEEE Transactions on Neural Networks*, 12:987-997, 2001.