Multivariate Functional Linear Discriminant Analysis Based on Pairwise Pseudo-Likelihood Modeling Combined with Splines

Kristien Wouters[1], José Cortiñas Abrahantes[1], Geert Molenberghs[1], Helena Geys[1,2], Abdellah Ahnaou[2], Pim Drinkenburg[2] and Luc Bijnens[2]

[1] Universiteit Hasselt, Center for Statistics, Belgium;
[2] Johnson & Johnson Pharmaceutical Research and Development, Belgium

In classification problems where the predictor variables are longitudinal, classical linear discriminant analysis (LDA) is not appropriate, since the correlation between different time points will be ignored. For this particular case, where only one longitudinal predictor variable is to be used, functional linear discriminant analysis as proposed by James and Hastie [1] can be applied. They proposed a method for generalizing linear discriminant analysis to functional data, which possesses all the usual LDA tools, including a low-dimensional graphical summary of the data, and classification of new curves.
In this paper, we propose an extension of this method for the case where several longitudinal profiles are recorded for the same individual. When dealing with more than one longitudinal variable, we have to take the correlation between these variables into account in our classification. A full multivariate model would be the natural choice, but given the complexity of the data, computational issues are commonly present during such modeling exercise. Therefore, we are proposing to use a pairwise pseudo-likelihood modeling approach [2] combined with smoothing techniques such as splines, to model the multivariate longitudinal characteristics. While the computational issues are overcome by this modeling approach, others arise, since the fitted covariance matrix obtained from the modeling approach is not always positive definite. To obtain a classification rule, we propose to calculate the average of the distance between a new observation and the classes in the training dataset for each pair of variables.
The performance of this procedure is established through simulation studies, using different number of classes, covariates and observations. The procedure is also applied to a real-life dataset where psychotropic drugs need to be classified according to their main indication in psychiatry based on electro-encephalograph (EEG) data [3], which can be used to determine sleep-wake architecture, when carried out in conjunction with movement monitoring and a so-called electromyogram (EMG) recording muscle activity. This clearly defines states of vigilance (in particular 6 sleeping stages) that can be separated out and used to classify psychotropic agents. In this data we deal with 6 longitudinal variables to be used to separate out 5 psychotropic drug classes. The performance is evaluated based on the classification error using a train-test setting, indicating a good performance of the procedure.

References

[1] James GM, Hastie TJ (2001) Functional Linear Discriminant Analysis for Irregularly Sampled Curves. Journal of the Royal Statistical Society. Series B 63(3):533-550

[2] Fieuws S, Verbeke G (2006) Pairwise Fitting of Mixed Models for the Joint Modeling of Multivariate Longitudinal Profiles. Biometrics 62:424-431

[3] Wouters K, Ahnaou A, Cortiñas J, Molenberghs G, Geys H, Bijnens L, Drinkenburg WHIM (2007) Psychotropic drug classification based on sleep-wake behaviour of rats. Journal of the Royal Statistical Society. Series C 56(2):223-234