Some topics on linear functional data analysis

Gilbert Saporta
Chaire de Statistique Appliquée & CEDRIC
Conservatoire National des Arts et Métiers
292 rue Saint Martin, F- 75141 Paris cedex 03
gilbert.saporta@cnam.fr

In a regression problem where we have a single response variable Y, the case of an infinite (not countable) set of predictors occurs when the predictors are functions or curves *ie* X_t with $t \in [0;T]$. Supervised classification into two categories by means of Fisher's linear discriminant function may be viewed as a special case where the dependent variable has 2 values.
Linear methods look for a predictor which may be expressed as an integral sum (Fisher, 1924):

$$\hat{Y} = \int_0^T X_t \beta(t) dt$$

When *t* takes continuously its values in an interval [0;T], multicollinearity leads to inconsistent estimation of the parameters.
In order to solve this problem, we focus here on linear methods based on an orthogonal decomposition of the predictors leading to dimensionality reduction. Components derived from the Karhunen-Loeve decomposition give, for functional data, the equivalent of principal components regression (PCR). However PLS (partial least squares) performs better than PCR, since principal components are obtained irrespective of the response (Preda, Saporta, 2005 and Preda & al., 2007).
It is often of practical interest to anticipate the prediction of the response earlier than the end of the observed process. A method based on the use of ROC curves and AUC is presented: For real time applications we determine an optimal time $s < T$ giving a prediction based on [0;s] almost as good as the prediction using all data (Costanzo & al, 2006) by using a bootstrap test for the AUC criterion.

References

Costanzo D., Preda C., Saporta G. (2006) Anticipated prediction in discriminant analysis on functional data for binary response . In *COMPSTAT2006*, p. 821-828, Physica-Verlag

Fisher R.A. (1924) The Influence of Rainfall on the Yield of Wheat at Rothamsted. *Philosophical Transactions of the Royal Society*, B: 213: 89-142

Preda C. , Saporta G. (2005) PLS regression on a stochastic process, *Computational Statistics and Data Analysis*, 48, 149-158.

Preda C., Saporta G., Lévéder C., (2007) PLS classification of functional data, *Computational Statistics*, 22(2), 223-235