CATEGORICAL DISCRIMINANT ANALYSIS BY SMOOTHING MULTIPLE CORRESPONDENCE ANALYSIS (Abstract)

The basic supposition of the method is that the categorical variables come from underlying ones, mixtures of normals, discretized in permuted intervals at the marginals. The effort was directed in reproducing the more exactly possible underlying continuous probability distribution and later to apply a methodology of discrimination with continuous variables. In order to obtain this reconstruction, two phases were developed: In first, a multiple correspondence analysis, properly adapted to the discriminant objective, looked for quantifications, and secondly, a procedure of smoothing based on algorithm EM, complemented the reproduction applying a dispersion around these values. In Chapter 1 the basic definitions of the discriminant analysis are analyzed and the existing methods had been revised with the mentioned objective. The second and third chapter make the equivalent with the correspondence analysis and the methods of smoothing (Kernel and EM) like elements to combine to obtain the reconstruction. Chapter 4 makes the methodologic proposal and a generalization of Lancaster's theorem is demonstrated for its mathematical foundation. Finally, in chapter 5, the results with simulated and real data are discussed, reaching the following conclusions: The test of the method with underlying normal simulated data using a model with common variance by class was positive since the method improved the other procedures which it has been compared with. On the other hand if the multinormal supposition of an underlying variable is interpreted like the end of a research when finally we are able to separate the significants factors affecting the average whereas the nonsignificant ones determine a gaussian dispersion is not surprising that a method based on these premises obtains good practical results, as it happens in real complex examples analyzed.