8

Functional Nonparametric Supervised Classification

This chapter presents a nonparametric kernel method for discriminating functional data. Because theoretical advances are easily derived from those obtained in the regression setting (see Section 8.5), this chapter emphasizes applied features. The method is described in Section 8.2, the computational issues are discussed in Section 8.3 and two case studies are reported in Section 8.4. This chapter ends with some comments and bibliographical notes.

8.1 Introduction and Problematic

Supervised classification or discrimination of functional data corresponds to the situation when we observe a f.r.v. $\boldsymbol{\mathcal{X}}$ and a categorical response Y which gives the class membership of each functional object. As illustration, you can refer to the speech recognition data in Section 2.2. The log-periodograms are the observations of the f.r.v. and the class membership is defined by their corresponding phonemes. The main aim in such a setting is to give reasonable answers to the following questions: given a new functional data, can we predict its class membership? Are we able to provide a consistent rule for assigning each functional object to some homogeneous group? What do we mean by a homogeneous group? How can we measure the performance of such a classification rule?

Before going on, is the classical linear discriminant analysis operational in such a setting? The answer is no because it is well known that a large number of predictors relative to the sample size and/or highly correlated predictors (which is the case when we consider functional data) lead to a degenerated within-class covariance matrix. In this functional context, the linear discrimination analysis fails. Therefore, alternative methods have been developed.

The next Section describes an alternative methodology for building a nonparametric classification rule. This is done through a proximity measure between the functional objects and a kernel estimator of the posterior probabilities derived from the one introduced in the functional nonparametric pre-

diction context. Section 8.3 focuses on practical aspects by giving a simple way to automatically choose both the smoothing parameter introduced in the kernel estimator and the one needed for the proximity measure. To illustrate, Section 8.4 proposes applications of such a functional nonparametric method to curves discrimination; our procedure is applied to the chemometric and speech recognition data. Section 8.5 gives some theoretical properties of our kernel estimator which are easily deduced from the asymptotic behaviour of the kernel estimator in the functional nonparametric regression setting (see for more details Sections 6.2.1 and 6.3.1). The last Section is devoted to the state of the art in this area and the bibliography therein.

8.2 Method

Let $(\boldsymbol{\mathcal{X}}_i, Y_i)_{i=1,\dots,n}$ be a sample of n independent pairs, identically distributed as $(\boldsymbol{\mathcal{X}}, Y)$ and valued in $E \times \overline{G} = \{1, \dots, G\}$, where (E, d) is a semi-metric vector space (i.e. $\boldsymbol{\mathcal{X}}$ is a f.r.v. and d a semi-metric). In practical situations, we will use the notation (χ_i, y_i) for the observation of the pair $(\boldsymbol{\mathcal{X}}_i, Y_i)$, for all i varying form 1 to n. To clarify the situation, you can keep in mind the speech recognition example (Section 2.2): the $\boldsymbol{\mathcal{X}}_i$'s are the log-periodograms whereas the y_i's are the corresponding classes of phoneme ($G = 5$).

General classification rule (Bayes rule). Given a functional object χ in E, the purpose is to estimate the G posterior probabilities

$$p_g(\chi) \;=\; P\left(Y = g | \boldsymbol{\mathcal{X}} = \chi\right), \quad g \in \overline{G}.$$

Once the G probabilities are estimated $(\widehat{p}_1(\chi), \dots, \widehat{p}_G(\chi))$, the classification rule consists of assigning an incoming functional observation χ to the class with highest estimated posterior probability:

$$\widehat{y}(\chi) \;=\; \arg\max_{g \in \overline{G}} \widehat{p}_g(\chi).$$

This classification rule is also called *Bayes rule*. In order to make precise our functional discriminant method, what remains is to build a suitable kernel estimator.

Kernel estimator of posterior probabilities. Before defining our kernel-type estimator of the posterior probabilities, we remark that

$$p_g(\chi) \;=\; \mathbb{E}\left(1_{[Y=g]} | \boldsymbol{\mathcal{X}} = \chi\right),$$

with $1_{[Y=g]}$ equals to 1 if $Y = g$ and 0 elsewhere. In this way, it is clear that the posterior probabilities can be expressed in terms of conditional expectations.