Classification in postural style

Antoine Chambaz, Christophe Denis

HAL Id: hal-00576070

https://hal.archives-ouvertes.fr/hal-00576070

Submitted on 11 Mar 2011

Classification in postural style

Antoine Chambaz and Christophe Denis*
MAP5, Université Paris Descartes and CNRS

March 11, 2011

Abstract

This article contributes to the search for a notion of *postural style*, focusing on the issue of classifying subjects in terms of how they maintain posture. Longer term, the hope is to make it possible to determine on a case by case basis which sensorial information is prevalent in postural control, and to improve/adapt protocols for functional rehabilitation among those who show deficits in maintaining posture, typically seniors. Here, we specifically tackle the statistical problem of classifying subjects sampled from a two-class population. Each subject (enrolled in a cohort of 54 participants) undergoes four experimental protocols which are designed to evaluate potential deficits in maintaining posture. This results in four complex trajectories, from which we decide to extract four small-dimensional summary measures. Because undergoing several protocols is at least unpleasant, and sometimes painful, we try to limit the number of protocols needed for the classification. Therefore, we first rank the protocols by decreasing order of relevance, then we derive four plug-in classifiers which involve the best (*i.e.*, more informative), the two best, the three best and all four protocols. This two-step procedure relies on the cutting-edge methodologies of targeted maximum likelihood learning (a novel methodology for robust and efficient estimation and testing) and super-learning (a machine learning procedure for aggregating various estimation procedures into a single better estimation procedure). A simulation study is carried out. The performances of the procedure applied to the real dataset (and evaluated by the leave-one-out rule) go as high as a 87% rate of correct classification (47 out of 54 subjects correctly classified), using only the best protocol.

1 Introduction

This article contributes to the search for a notion of *postural style*, focusing on the issue of classifying subjects in terms of how they maintain posture.

Posture is fundamental to all activities, including locomotion and prehension. Posture is the fruit of a dynamic analysis by the brain of visual, proprioceptive and vestibular informations. Proprioceptive information stems from the ability to sense the position, location, orientation and movement of the body and its parts. Vestibular information roughly relates to the sense of equilibrium. Naturally, every individual has developed his/her own preferences according to his/her sensorimotor experience. Sometimes, a sole kind of information (usually visual) is processed in all situations. Although this kind of processing may be efficient for maintaining posture in one's usual environment, it is likely not adapted to reacting to new or unexpected situations. Such situations may result in falling, the consequences of a fall being particularly bad in seniors. Longer term, the hope is to make it possible to determine on a case by case basis which sensorial information is prevalent in postural control, and to improve/adapt protocols for functional rehabilitation among those who show deficits in maintaining posture, typically seniors.

As in earlier studies (see [2, 5] and references therein), our approach to characterizing postural control involves the use of a force-platform. Subjects standing on a force-platform are exposed to different perturbations, following different experimental protocols (or simply *protocols* in the sequel). The force-platform records over time the center-of-pressure of each foot, that is "the position of the global ground reactions forces that accommodates the sway of the body" [10]. A protocol is divided into three phases: a first phase without perturbation, followed by a second phase with perturbation, followed itself by a last phase without perturbation. Different kind of perturbations are considered. They can be characterized either as visual, or proprioceptive or vestibular, depending on which sensorial system is perturbed.

We specifically tackle the statistical problem of classifying subjects sampled from a two-class population. The first class regroups subjects who do not show any deficit in postural control. The second class regroups hemiplegic subjects, who suffer from a proprioceptive deficit. Even though differentiating two subjects from the two groups is relatively easy by visual inspection, it is a much more delicate task when relying on some general baseline covariates and the trajectories provided by a force-platform. Furthermore, since undergoing several protocols is at least unpleasant, and sometimes painful (some sensitive subjects have to lie down for 15 minutes in order to recover from dizziness after a series of protocols), we also try to limit the number of protocols used for classifying.

Our classification procedure relies on cutting-edge statistical methodologies. In particular, we come up with a nice preliminary ranking of the four protocols (in view of how much we can learn from them on postural control) which involves the targeted maximum likelihood methodology [16, 12], a novel statistical procedure for robust and efficient estimation and testing. The targeted maximum likelihood methodology relies itself on the super-learning procedure, a machine learning methodology for aggregating various estimation procedures (or simply *estimators*) into a single better estimation procedure/estimator [15, 12]. In addition to being a key element of the targeted maximum likelihood ranking of the protocols, the super-learning procedure plays also a crucial role in the construction of our classification procedure itself.

We show that it is possible to achieve a 87% rate of correct classification (47 out of 54 subjects correctly classified; the performance is evaluated by the leave-one-out rule), using only the more informative protocol. Our classification procedure is easy to generalize (we actually provide an example of generalization), so we reasonably hope that even better results are within reach (especially considering that more data should soon augment our small dataset). The interest of the article goes beyond the specific application. It nicely illustrates the versatility and power of the targeted maximum likelihood and super-learning methodologies. It also shows that retrieving and comparing small-dimensional summary measures from complex trajectories may be convenient to classify them.

The article is organized as follows. In Section 2, we describe the dataset which is at the core of the study. The classification procedure is formally presented in Section 3, and its performances, evaluated by simulations, are discussed in Section 4. We report in Section 5 the results obtained by applying the latter classification procedure to the real dataset. We relegate to the Appendix a self-contained presentation of the super-learning procedure as it is used here, and the description of an estimation procedure/estimator that will play a great role in the super-learning procedure applied to the construction of our classification procedure.

2 Data description

The dataset, collected at the Center for the study of sensorimotor functioning (CESEM) of the University Paris Descartes, is described in Section 2.1. We motivate the introduction of a summarized version of each observation, and present its construction in Section 2.2.

2.1 Original dataset

Each subject undergoes four protocols that are designed to evaluate potential deficits in maintaining posture. The specifics of the latter protocols are presented in Table 1. Protocols 1 and 2 respectively perturb the processing of visual data and proprioceptive information by the brain. Protocol 3 cumulates both perturbations. Protocol 4 relies on perturbing the processing of vestibular information by the brain through a visual stimulation.

A total of $n = 54$ subjects are enrolled. For each of them, the age, gender, laterality (the preference that most humans show for one side of their body over the other), height and weight are collected. Among the 54 subjects, 22 are hemiplegic (due to a cerebrovascular accident), and therefore suffer from a proprioceptive deficit in postural control. Initial medical examinations concluded that the 32 other subjects show no pronounced deficits in postural control. We will refer to those subjects as *normal subjects*.

protocol	1st phase (0→15s)	2nd phase (15→50s)	3rd phase (50→70s)
1		eyes closed	
2	no perturbation	muscular stimulation	no perturbation
3		eyes closed muscular stimulation	
4		optokinetic stimulation	

Table 1: Specifics of the four protocols designed to evaluate potential deficits in postural control. A protocol is divided into three phases: a first phase without perturbation of the posture is followed by a second phase with perturbations, which is itself followed by a last phase without perturbation. Different kind of perturbations are considered. They can be characterized either as visual (closing the eyes), or proprioceptive (muscular stimulation) or vestibular (optokinetic stimulation), depending on which sensorial information is perturbed.

For each protocol, the center-of-pressure of each foot is recorded over time. Thus each protocol results in a trajectory $(X_t)_{t \in T} = (L_t, R_t)_{t \in T}$ where $L_t = (L_t^1, L_t^2) \in \mathbb{R}^2$ (respectively, $R_t = (R_t^1, R_t^2)$) gives the position of the center-of-pressure of the left (respectively, right) foot on the force-platform at time t, for each t in $T = \{k\delta : 1 \leq k \leq 2800\}$ where the time-step $\delta = 0.025$ seconds (the protocol lasts 70 seconds). We represent in Figure 1 two such trajectories $(X_t)_{t \in T}$ associated with a normal subject and a hemiplegic subject, both undergoing the third protocol (see Table 1). Note that we do not take into account the first few seconds of the recording that a generic subject needs to reach a stationary behavior.

Figure 1 confirms the intuition that the structure of a generic trajectory $(X_t)_{t \in T}$ is complicated, and that a mere visual inspection is, at least on this example, of little help for differentiating the normal and hemiplegic subjects. Although several articles investigate how to model and use such trajectories directly [2, 5], we rather choose to rely on a summary measure of $(X_t)_{t \in T}$ instead of relying on $(X_t)_{t \in T}$ itself.

2.2 The making of the summary measure

The summary measure that we construct is actually a summary measure of a one-dimensional trajectory $(C_t)_{t \in T}$ that we initially derive from $(X_t)_{t \in T}$. First, we introduce the trajectory of barycenters, $(B_t)_{t \in T} = (\frac{1}{2}(L_t + R_t))_{t \in T}$. Second, we evaluate a reference position b which is defined as the componentwise median value of $(B_t)_{t \in T \cap [0,15]}$ (that is, the median value over the first phase of the protocol). Third, we set $C_t = \|B_t - b\|_2$ for all $t \in T$, the Euclidean distance between B_t and the reference position b, which provides a relevant description of the sway of the body during the course of the protocol. We plot in Figure 2 two examples of $(C_t)_{t \in T}$ corresponding to two different protocols undergone by a hemiplegic subject.

Now, it is arguably in the neighborhood (in time) of the beginning and ending of the second phase that the most characteristic features of a trajectory should be sought. As an illustration, it is striking that one recovers easily the beginning and ending times of the second phase of the
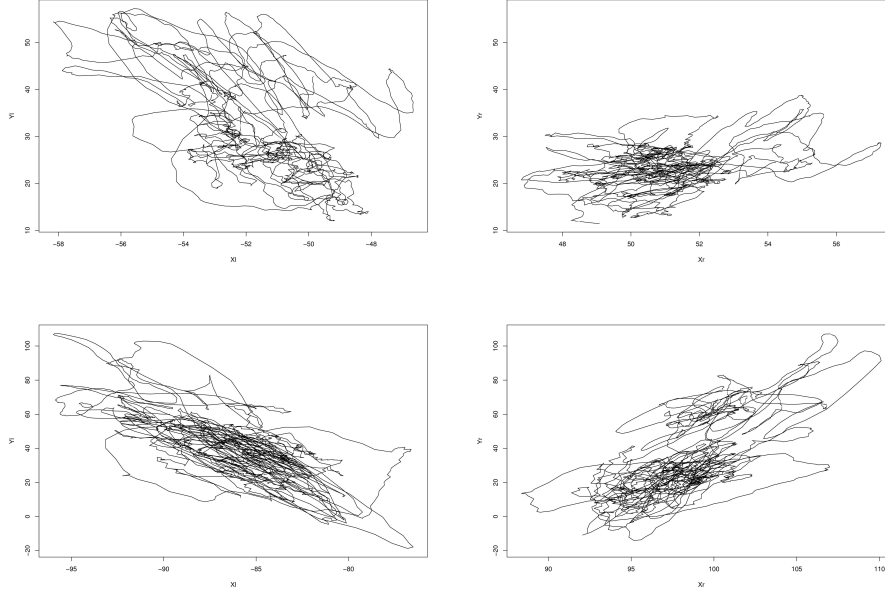
3

Figure 1: Sequences $t \mapsto L_t$ (left) and $t \mapsto R_t$ (right) of positions of the center-of-pressure over T of both feet on the force-platform, associated with a normal subject (top) and a hemiplegic subject (bottom), who undergo the third protocol (see Table 1).

third protocol from the right-hand plot in Figure 2, but not for the protocol 1 from the left-hand plot of the same figure. Therefore, we decide to focus on the following finite-dimensional summary measure of $(X_t)_{t \in T}$ (through $(C_t)_{t \in T}$):

$$Y = (\bar{C}_1^+ - \bar{C}_1^-, \bar{C}_2^- - \bar{C}_1^+, \bar{C}_2^+ - \bar{C}_2^-) \tag{1}$$

where

$$\bar{C}_1^- = \frac{\delta}{5} \sum_{t \in T \cap [10,15[} C_t, \qquad \bar{C}_1^+ = \frac{\delta}{5} \sum_{t \in T \cap]15,20]} C_t,$$

$$\bar{C}_2^- = \frac{\delta}{5} \sum_{t \in T \cap [45,50[} C_t, \qquad \bar{C}_2^+ = \frac{\delta}{5} \sum_{t \in T \cap]50,55]} C_t$$

are the averages of C_t computed over the intervals $[10, 15[,]15, 20], [45, 50[$ and $]50, 55]$ (that is over the last/first 5 seconds before/after the beginning/ending of the second phase of the protocol of interest). We arbitrarily choose this 5-second threshold. Note that it is not necessary to consider the remaining differences $\bar{C}_2^- - \bar{C}_1^- = Y_2 + Y_1, \bar{C}_2^+ - \bar{C}_1^- = Y_3 + Y_2, \bar{C}_2^+ - \bar{C}_1^+ = Y_1 + Y_2 + Y_3$ since they all are linear combinations of the components of Y. We refer to Figure 3 for a visual representation of the definition of the summary measure Y.

3 Classification procedure

We describe hereafter our two-step classification procedure. We formally introduce the statistical framework that we consider in Section 3.1. The first step of the classification procedure consists of ranking the protocols from the most to the less informative with respect to some criterion, see Section 3.2. The second step consists of the classification itself, see Section 3.3.
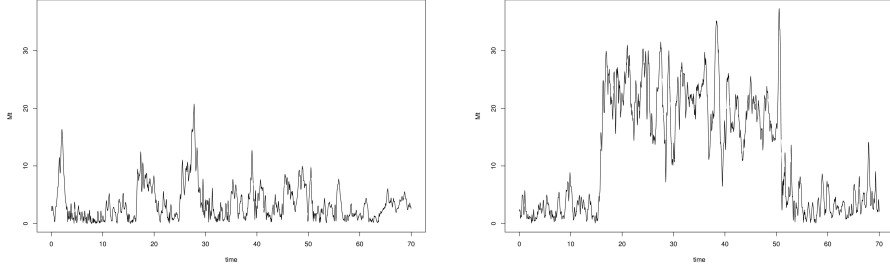
Figure 2: Representing the trajectories $t \mapsto C_t$ over T which correspond to two different protocols undergone by a hemiplegic subject (protocol 1 on the left, protocol 3 on the right).
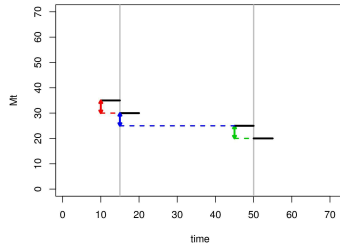

Figure 3: Visual representation of the definition of the finite-dimensional summary measure Y of $(X_t)_{t \in T}$. The four horizontal segments (solid lines) represent, from left to right, the averages $\bar{C}_1^-$, $\bar{C}_1^+$, $\bar{C}_2^-$, $\bar{C}_2^+$ of $(C_t)_{t \in T}$ over the intervals $[10, 15[$, $]15, 20]$, $[45, 50[$, $]50, 55]$. The three vertical segments (solid lines ending by an arrow) represent, from top to bottom, the components Y_1, Y_2 and Y_3 of Y. Two additional vertical lines indicate the beginning and ending of the second phase of the considered protocol.

3.1 Statistical framework

The observed data structure O writes as $O = (W, A, Y^1, Y^2, Y^3, Y^4)$, where

- $W \in \mathbb{R} \times \{0, 1\}^2 \times \mathbb{R}^2$ is the vector of baseline covariates (corresponding to initial age, gender, laterality, height and weight, see Section 2.1);

- $A \in \{0, 1\}$ indicates the subject's class (with convention $A = 1$ for hemiplegic subjects and $A = 0$ for normal subjects);

- for each $j \in \{1, 2, 3, 4\}$, $Y^j \in \mathbb{R}^3$ is the summary measure (as defined in (1)) associated with the jth protocol.

We denote by P_0 the true distribution of O. Since we do not know much about P_0, we simply see it as an element of the non-parametric set $\mathcal{M}$ of all possible distributions of O.

We need a criterion to rank the four protocols from the most to the less informative in view of the subject's class. To this end, we introduce the functional $\Psi : \mathcal{M} \to \mathbb{R}^{12}$ such that, for any $P \in \mathcal{M}$, $\Psi(P) = (\Psi^j(P))_{1 \leq j \leq 4}$ where

$$\Psi^j(P) = \left(E_P \left\{ E_P[Y_i^j | A = 1, W] - E_P[Y_i^j | A = 0, W] \right\} \right)_{1 \leq i \leq 3}.$$

The component $\Psi_i^j(P)$ is known in the literature as the variable importance measure of A on the summary measure Y_i^j controlling for W [12]. Under causal assumptions, it can be interpreted as the effect of A on Y_i^j. More generally, we are interested in $\Psi_i^j(P_0)$ because the further it is from zero,

the more knowledge on A we expect to gain from the observation of W and the summary measure Y_i^j (*i.e.*, by comparing the averages of $(C_t)_{t \in T}$ computed over the time intervals corresponding to index i, see Section 2.2). For instance, say that $\Psi_1^2(P_0) > 0$: this means that (in P_0-average), the variation in mean of the mean postures $\bar{C}_1^-$ and $\bar{C}_1^+$ of a hemiplegic subject computed before and after the beginning of the muscular perturbation is larger than that of a normal subject. In words, the postural control of a hemiplegic subject is more affected by the beginning of the muscular perturbation than the postural control of a normal subject.

3.2   Targeted maximum likelihood ranking of the protocols

Our ranking of the four protocols relies on testing the null hypotheses

$$\text{``}\Psi_i^j(P_0) = 0\text{''}, \quad (i,j) \in \{1,2,3\} \times \{1,2,3,4\},$$

against their two-sided alternatives. Heuristically, rejecting "$\Psi_i^j(P_0) = 0$" tells us that the value of the ith coordinate of the summary measure Y^j provides helpful information for the sake of determining whether $A = 0$ or $A = 1$.

We consider tests based on the targeted maximum likelihood methodology [16, 12]. It is a novel procedure for robust and efficient estimation and testing. Because presenting a self-contained introduction to the methodology would significantly lengthen the article, we only provide below a *very succinct* description of it. The targeted maximum likelihood methodology relies itself on the super-learning procedure, a machine learning methodology for aggregating various estimation procedures (or simply *estimators*) into a single better estimation procedure/estimator [15, 12], based on the cross-validation principle. Since super-learning also plays a crucial role in our classification procedure (see Section 3.3), and because it is possible to present a relatively short self-contained introduction to the construction of a super-learner, we propose such an introduction in Section A.1.

Let $O_{(1)}, \ldots, O_{(n)}$ be n independent copies of the observed data structure O. For every $(i,j) \in \{1,2,3\} \times \{1,2,3,4\}$, we compute the targeted maximum likelihood estimator (TMLE) $\Psi_{i,n}^j$ of $\Psi_i^j(P_0)$ based on $O_{(1)}, \ldots, O_{(n)}$ and an estimator $\sigma_{i,n}^j$ of its asymptotic standard deviation $\sigma_i^j(P_0)$. The methodology applies because Ψ_i^j is a "smooth" parameter. It notably involves the super-learning of the conditional means $Q_i^j(P_0)(A,W) = E_{P_0}(Y_i^j|A,W)$ and of the conditional distribution $g(P_0)(A|W) = P_0(A|W)$ (the collection of estimators aggregated by super-learning is given in Sections A.1). Under some regularity conditions, the estimator $\Psi_{i,n}^j$ of $\Psi_i^j(P_0)$ is consistent when either $Q_i^j(P_0)$ or $g(P_0)$ is consistently estimated, and it satisfies a central limit theorem. In addition, if $g(P_0)$ is consistently estimated by a maximum-likelihood based estimator, then $\sigma_{i,n}^j$ is a conservative estimator of $\sigma_i^j(P_0)$. This finally yields, for every $(i,j) \in \{1,2,3\} \times \{1,2,3,4\}$, a t-statistics $T_{i,n}^j = \frac{\sqrt{n}\Psi_{i,n}^j}{\sigma_{i,n}^j}$.

Now, we rank the four protocols by comparing the 3-dimensional vectors of test statistics $(T_{1,n}^j, T_{2,n}^j, T_{3,n}^j)$ for $1 \leq j \leq 4$. Several criteria for comparing the vectors were considered. They all relied on the fact that the larger is $|T_{i,n}^j|$ the less likely the null "$\Psi_i^j(P_0) = 0$" is true. Since the results were only slightly affected by the criterion, we focus here on a single one. Thus, we decide that protocol j is more informative than protocol j' if

$$\sum_{i=1}^{3}(T_{i,n}^{j'})^2 < \sum_{i=1}^{3}(T_{i,n}^j)^2.$$

This rule is motivated by the fact that, if $\sigma_{1,n}^j, \sigma_{2,n}^j, \sigma_{3,n}^j$ are consistent estimators of $\sigma_1^j(P_0), \sigma_2^j(P_0), \sigma_3^j(P_0)$, then $\sum_{i=1}^{3}(T_{i,n}^j)^2$ asymptotically follows the $\chi^2(3)$ distribution under H_0^j: "$\Psi^j(P_0) = 0$".

By definition of the observed data structure and by construction of the TMLE procedure, this rule yields almost surely a final ranking of the four protocols from the more to the less informative for the sake of determining whether $A = 0$ or $A = 1$.

3.3 Classifying a new subject

We now tackle the main goal of this article: Building a classifier ϕ for determining whether $A = 0$ or $A = 1$ based on the baseline covariates W and summary measures (Y^1, Y^2, Y^3, Y^4). In order to study the influence of the ranking on the classification, we actually build four different classifiers $\phi^1, \phi^2, \phi^3, \phi^4$ which respectively use only the best (more informative) protocol, the two best, the three best, and all four protocols. So ϕ^j is a function of W and of j among the four vectors Y^1, Y^2, Y^3, Y^4.

Say that $\mathcal{J} \subset \{1, 2, 3, 4\}$ has J elements. First, we build an estimator $h_n^J(W, Y^j, j \in \mathcal{J})$ of $P_0(A = 1 | W, Y^j, j \in \mathcal{J})$ based on $O_{(1)}, \ldots, O_{(n)}$, relying again on the super-learning methodology (the self-contained introduction to the construction of a super-learner of Section A.1 is augmented by the description of the collection of estimators involved in the super-learning). Second, we define

$$\phi^J(W, Y^j, j \in \mathcal{J}) = \mathbf{1}\{h_n(W, Y^j, j \in \mathcal{J}) \geq \tfrac{1}{2}\}$$

and decide to classify a new subject with information $(W, Y^j, j \in \mathcal{J})$ into the group of hemiplegic subjects if $\phi^J(W, Y^j, j \in \mathcal{J}) = 1$ or into the group of normal subjects otherwise.

Thus, the classifier ϕ^J relies on a plug-in rule, in the sense that the Bayes decision rule $\mathbf{1}\{P_0(A = 1 | W, Y^j, j \in \mathcal{J}) \geq \tfrac{1}{2}\}$ is mimicked by the empirical version where one substitutes an estimator of $P_0(A = 1 | W, Y^j, j \in \mathcal{J})$ for the latter regression function. Such classifiers can converge with fast rates under a complexity assumption on the regression function and the so-called margin condition [1].

4 Simulation study

In this section, we carry out and report the results of a simulation study of the performances of the classification procedure described in Section 3. The details of the simulation scheme are presented in Section 4.1, and the results are reported and evaluated in Section 4.2.

4.1 Simulation scheme

Instead of simulating (W, A) and the four complex trajectories $(X_t^1)_{t \in T}$, $(X_t^2)_{t \in T}$, $(X_t^3)_{t \in T}$, $(X_t^4)_{t \in T}$ associated with four fictitious protocols (see Section 2.1), we generate directly (W, A) and the summary measures Y^1, Y^2, Y^3, Y^4 that one would have derived from $(X_t^1)_{t \in T}$, $(X_t^2)_{t \in T}$, $(X_t^3)_{t \in T}$, $(X_t^4)_{t \in T}$ (see Section 2.2). Three different scenarios/probability distributions P_0^1, P_0^2, P_0^3 are considered. They only differ from each other with respect to the conditional distributions $g(P_0^1)$, $g(P_0^2)$, $g(P_0^3)$ (see Table 2 for their characterization).

For each $k = 1, 2, 3$, a generic observed data structure $O = (W, A, Y^1, Y^2, Y^3, Y^4)$ drawn from P_0^k meets the following constraints:

1. W is drawn from a slightly perturbed version of the empirical distribution of W as obtained from the original dataset (the same for all $k = 1, 2, 3$);

2. conditionally on W, A is drawn from $g(P_0^k)(\cdot | W)$;

3. conditionally on (A, W) and for each $(i, j) \in \{1, 2, 3\} \times \{1, 2, 3, 4\}$, Y_i^j is drawn from the Gaussian distribution with mean $Q_i^j(A, W)$ (the same for all $k = 1, 2, 3$; see Table 3 for the definition of the conditional means) and common standard deviation $\sigma \in \{0.5, 1\}$.

Although that may not be clear when looking at Table 2, the difficulty of the classification problem should vary from one scenario to the other. When using the first conditional distribution $g(P_0^1)$, the conditional probability of $A = 1$ given W is concentrated around $\frac{1}{2}$, as seen in Figure 4 (solid line), with $P_0^1(g(P_0^1)(1|W) \in [0.48, 0.54]) \simeq 1$. In words, the covariate provides little information for predicting the class A. On the contrary, estimating $g(P_0^1)$ from the data is easy since $\text{logit } g(P_0^1)(A = 1|W)$ is a simple linear function of W. The conditional probabilities of $A = 1$ given W under $g(P_0^2)$ and $g(P_0^3)$ are less concentrated around $\frac{1}{2}$, as seen in Figure 4 (dashed and

dotted lines, respectively). Thus, the covariates may provide valuable information for predicting the class. But this time, logit $g(P_0^2)$ and logit $g(P_0^3)$ are tricky functions of W.

Likewise, the family of conditional means $Q_i^j(A, W)$ of Y_i^j given (A, W) that we use in the simulation scheme is meant to cover a variety of situations with regard to how difficult it is to estimate each of them and how much they tell about the class prediction. Instead of representing the latter conditional means, we find it more relevant to provide the reader with the values (computed by Monte-Carlo simulations) of

$$S^j(P_0^k) = \sum_{i=1}^{3} \left(\frac{\Psi_i^j(P_0^k)}{\sigma_i^j(P_0^k)} \right)^2$$

for $(j, k) \in \{1, 2, 3, 4\} \times \{1, 2, 3\}$ and $\sigma \in \{0.5, 1\}$, see Table 4. Indeed, $nS^j(P_0^k)$ should be interpreted as a theoretical counterpart to the criterion $\sum_{i=1}^{3}(T_{i,n}^j)^2$. In particular, we derive from Table 4 the theoretical ranking of the protocols: for every scenario P_0^k and $\sigma \in \{0.5, 1\}$, the protocols ranked by decreasing order of informativeness are protocols 3, 2, 1, 4.
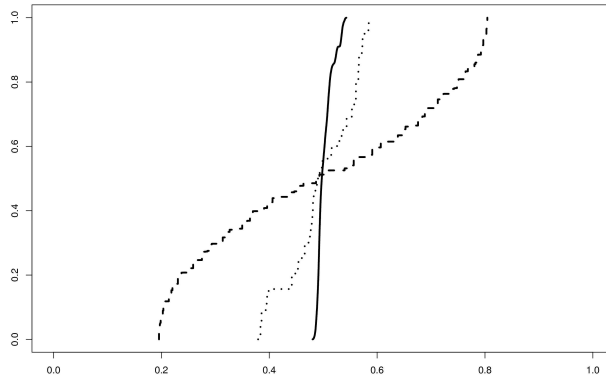

Figure 4: Visual representation of the three conditional distributions considered in the simulation scheme. We plot the empirical cumulative distribution functions of $\{g^k(A = 1|W_{(\ell)}) : \ell = 1, \ldots, n\}$ for $k = 1$ (solid line), $k = 2$ (dashed line) and $k = 3$ (dotted line), where $W_{(1)}, \ldots, W_{(L)}$ are independent copies of W drawn from the marginal distribution of W under P_0^k (which does not depend on k), and $L = 10^5$.

4.2 Leave-one-out evaluation of the performances of the classification procedure

We rely on the leave-one-out rule to evaluate the performances of the classification procedure. Specifically, we repeat independently $B = 100$ times the following steps for $k = 1, 2, 3$:

1. Draw independently $O_{(1,b)}, \ldots, O_{(n,b)}$ from P_0^k, with $n = 54$; we denote by $A_{(\ell,b)}$ the group membership indicator associated with $O_{(\ell,b)}$, and by $O'_{(\ell,b)}$ the observed data structure $O_{(\ell,b)}$ *deprived of $A_{(\ell,b)}$.*

2. For each $\ell \in \{1, \ldots, n\}$,

   (a) set $\mathcal{S}_{(\ell,b)} = \{O_{(\ell',b)} : \ell' \neq \ell, \ell' \leq n\}$;

   (b) based on $\mathcal{S}_{(\ell,b)}$, rank the protocols (see Section 3.2) then build four different classifiers $\phi_{(\ell,b)}^1$, $\phi_{(\ell,b)}^2$, $\phi_{(\ell,b)}^3$ and $\phi_{(\ell,b)}^4$ (see Section 3.3), which respectively use only the best (more informative), the two best, the three best and all four protocols (thus $\phi_{(\ell,b)}^J$ is a function of the covariate W and of J among the four vectors Y^1, Y^2, Y^3, Y^4);

8

$$\begin{aligned}
&\text{scenario 1: logit } g(P_0^1)(A=1|W) = \frac{W_1}{50} + \frac{W_2}{50} - \frac{W_3}{10} - \frac{W_4}{2000} + W_5 \\
&\text{scenario 2: logit } g(P_0^2)(A=1|W) = \cos(W_1 + W_5) + \sin(W_1 + W_5) \\
&\text{scenario 3: logit } g(P_0^3)(A=1|W) = \lfloor 10\cos(W_1 + W_3) \rfloor \\
&\quad + \sqrt{5\cos(W_1 + W_3) - \lfloor 5\cos(W_1 + W_3) \rfloor} \frac{\pi}{50} \sin(10\cos(W_1 + W_3))
\end{aligned}$$

Table 2: Characterization of the three conditional distributions $g(P_0^k)$, $k = 1, 2, 3$, considered in the simulation scheme.

fictitious protocol	conditional means
$j = 1$	$Q_1^1(A,W) = 2[A\sin(W_1+W_4) + (1-A)\cos(W_1+W_5)]$ $Q_2^1(A,W) = 3[(1-6A)X^5 - AX^4 + X^3 - (1-\frac{A}{2})X^2 + AX]$ where $X = \frac{(1-2A)W_5}{160} + \frac{A}{4}$ $Q_3^1(A,W) = A\tan(W_4) + (1-A)\tan(W_5 + W_1W_2)$
$j = 2$	$Q_1^2(A,W) = \frac{1}{120}[A + W_1 + W_2 + W_3 + W_5 + W_1W_2$ $+ (1-A)W_5 + W_2W_3W_4]$ $Q_2^2(A,W) = 5[A\sin(W_1+W_4) + (1-A)\cos(W_1+W_4)]$ $Q_3^2(A,W) = \frac{1}{20}[A(2W_1 + \frac{3}{2}W_3) + (1-A)W_5]$
$j = 3$	$Q_1^3(A,W) = A\log(2W_1 + \frac{3}{2}W_3) + (1-A)\log(W_5)$ $Q_2^3(A,W) = \frac{1}{45}(X+7)(X+2)(X-7)(X-3)$ where $X = \frac{W_4+W_5}{145+AW_1}$ $Q_3^3(A,W) = \pi[A\sin(X)(\lfloor 2X \rfloor + \sqrt{2X - \lfloor 2X \rfloor})$ $+ (1-A)\cos(X)(\lfloor 2X \rfloor + \sqrt{2X - \lfloor 2X \rfloor})]$ where $X = \cos(W_3 + W_4 + W_5)$
$j = 4$	$Q_1^4(A,W) = \frac{1}{100}(2X^3 + X^2 - X - 1)$ where $X = \frac{AW_2 + W_4 + W_5}{30}$ $Q_2^4(A,W) = \frac{1}{60}(A + W_1 + W_2 + W_3 + W_5)$ $Q_3^4(A,W) = \frac{1}{1000}[\frac{W_1W_3W_4}{3} + (1-A)(W_1 + W_3W_4) + AW_2W_5]$

Table 3: Conditional means $Q_i^j(A,W)$ of Y_i^j given (A,W) used in the three different scenarios of the simulation scheme.

fictitious protocol	scenario 1		scenario 2		scenario 3	
	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 1$
$j = 1$	0.14	0.04	0.11	0.03	0.14	0.04
$j = 2$	0.86	0.37	0.74	0.31	0.85	0.37
$j = 3$	2.94	1.12	2.49	0.93	2.90	1.11
$j = 4$	0.06	0.01	0.04	0.01	0.06	0.01

Table 4: Values of $S^j(P_0^k)$ for $(j,k) \in \{1,2,3,4\} \times \{1,2,3\}$ and $\sigma \in \{0.5, 1\}$. They notably teach us that, for every scenario P_0^k and $\sigma \in \{0.5, 1\}$, the protocols ranked by decreasing order of informativeness are protocols 3, 2, 1, 4.

(c) classify $O_{(\ell,b)}$ according to the four different classifications $\phi^1_{(\ell,b)}(O'_{(\ell,b)})$, $\phi^2_{(\ell,b)}(O'_{(\ell,b)})$, $\phi^3_{(\ell,b)}(O'_{(\ell,b)})$, $\phi^4_{(\ell,b)}(O'_{(\ell,b)})$.

3. Compute $\text{Perf}^J_b = \frac{1}{n}\sum_{\ell=1}^n \mathbf{1}\{A_{(\ell,b)} = \phi^J_{(\ell,b)}(O'_{(\ell,b)})\}$ for $J = 1,2,3,4$.

From this, we compute for each $J \in \{1,2,3,4\}$ the mean and standard deviation of the sample $(\text{Perf}^J_1,\ldots,\text{Perf}^J_B)$. There is no real need to report the values in a table, because they present a very clear pattern. First, all the standard deviations are approximately equal to 5%. Second, for every value of $\sigma \in \{0.5,1\}$, performance Perf^J actually depends only slightly on J (*i.e.*, on the number of protocols taken into account in the classification procedure), without any significant difference for $j = 1,2,3,4$. Third, the latter performances all equal approximately 80% when $\sigma = 1$, and increase to approximately 90% when $\sigma = 0.5$. This increase is the expected illustration of the fact the larger is the variability of the summary measures, the more difficult is the classification procedure. On the contrary, it is a little bit surprising that the conditional distributions $g(P_0^1), g(P_0^2), g(P_0^3)$ do not affect significantly the performances. Anecdotally, the estimated ranking of the protocols always coincide with the ranking that we derived from Table 4.

5 Application to the real dataset

We present here the results of the classification procedure of Section 3 applied to the real dataset. Thus, we first rank the protocols from the more to the less informative regarding postural control, see Section 5.1; then we construct the four classifiers and rely on the leave-one-out rule to evaluate their performances, see Section 5.2. A natural extension of the classification procedure is finally considered and applied in Section 5.3, yielding significantly better results.

5.1 Targeted maximum likelihood ranking of the protocols over the real dataset

It is a known medical fact that hemiplegic subjects are sensitive to muscular stimulations, and also that they tend to compensate for their proprioceptive deficit by developing a preference for visual information in order to maintain posture [3]. This suggests that protocols involving muscular and/or visual stimulations should rank high. What do the data tell us?

We derive and report in Table 5 the results of the ranking of the protocols using the entire dataset. Table 5 teaches us that the most informative protocol is protocol 3 (visual *and* muscular stimulations), and that the three next protocols ranked by decreasing order of informativeness are protocols 2 (muscular stimulation), 1 (visual stimulation), and 4 (optokinetic stimulation). Apparently, protocols 3 and 2 (which have in common that muscular stimulations are involved) are highly relevant for differentiating normal and hemiplegic subjects based on postural control data. On the contrary (and perhaps surprisingly, given the introductory remark), protocols 1 and 4 seem to provide significantly less information for the same purpose.

protocol	$j = 3$	$j = 2$	$j = 1$	$j = 4$
criterion $\sum_{i=1}^3 (T^j_{i,n})^2$	75.51	33.13	6.80	5.53

Table 5: Ranking the four protocols using the entire real dataset. We report the realizations of the criteria $\sum_{i=1}^3 (T^j_{i,n})^2$ obtained for protocols $j = 1,2,3,4$. These values teach us that the most informative protocol is protocol 3, and that the three next protocols ranked by decreasing order of informativeness are protocols 2, 1, and 4.

5.2 Classification procedures applied to the real dataset

In order to evaluate the performances of the classification procedure applied to the real dataset, we carry out steps 2a, 2b, 2c from the leave-one-out rule described in Section 4.2 where we

substitute the real dataset $O_{(1)}, \ldots, O_{(n)}$ for the simulated one. We actually do it *twice*. The first time, the super-learning methodology involves a large collection of estimators (see Section A.1). Then we justify resorting to a smaller collection (see Section A.1 again), hence a second round of performance evaluation. We report the results in Table 6, where the second and third rows respectively correspond to the first (larger collection) and second (smaller collection) round of performance evaluation.

Let's consider first the performances of the classification procedure relying on the larger collection. The proportion of subjects correctly classified (evaluated by the leave-one-out rule) equals only 70% (38 out of the 54 subjects are correctly classified) when the sole most informative protocol (*i.e.*, protocol 3) is exploited. This rate jumps to 80% (43 out of 54 subjects are correctly classified) when the two most informative protocols (*i.e.*, protocols 3 and 2) are exploited. Interestingly, including one or two of the remaining protocols decreases the performances.

	$J = 1$	$J = 2$	$J = 3$	$J = 4$
Perf^J (larger collection)	0.70 (38/54)	0.80 (43/54)	0.74 (40/54)	0.78 (42/54)
Perf^J (smaller collection)	0.74 (40/54)	0.81 (44/54)	0.78 (42/54)	0.85 (46/54)

Table 6: Leave-one-out performances Perf^J of the classification procedure using the real dataset. Performance Perf^J corresponds to the classifier based on J among the four vectors Y^1, Y^2, Y^3, Y^4 (those associated with the J more informative protocols) and either using all estimators (second row) or only two of them (third row) in the super-learner (see Section A.1 for details).

Now, the theoretical properties of the super-learning procedure [15, 12] are asymptotic, *i.e.* valid when the sample size n is large, which is arguably not the case in this study. Even though this is contradictory to the philosophy of the super-learning methodology, it is tempting to reduce the number of estimators involved in the super-learning. We therefore keep only two of them (see Section A.1 for the details), and run again steps 2a, 2b, 2c from the leave-one-out rule described in Section 4.2 where we substitute the real dataset $O_{(1)}, \ldots, O_{(n)}$ for the simulated one. Results are reported in Table 6 (third row). We obtain better performances: for each value of J (*i.e.*, each number of protocols taken into account in the classification procedure), the second classifier outperforms the first one. The best performance is achieved when all four protocols are used, yielding a rate of correct classification equal to 85% (46 out of the 54 subjects are correctly classified). This is encouraging, notably because one can reasonably expect that performances will be improved on when a larger cohort is available.

Yet, this is not the end of the story. We have built a general methodology that can be easily extended, for instance by enriching the small-dimensional summary measure derived from each complex trajectory. We explore the effects of such an extension in the next section.

5.3 Extension

Thus, let us enrich the small-dimensional summary measure initially defined in Section 2.2. Since it mainly involves *distances* from a reference point, the most natural extension is to add informations pertaining to *orientation*. Relying on polar coordinates of the trajectory $(B_t)_{t \in T}$ poses some technical issues. Instead, we propose to fit simple linear models $y(B_t) = vx(B_t) + u$ (where $x(B_t)$ and $y(B_t)$ are the abscisse and ordinate of B_t) based on the datasets $\{B_t : t \in T \cap [10, 15[\}$, $\{B_t : t \in T \cap [15, 20[\}$, $\{B_t : t \in T \cap [20, 45[\}$, $\{B_t : t \in T \cap [45, 50[\}$ and $\{B_t : t \in T \cap [50, 55[\}$, and to use the *slope estimates* as summary measures of an average orientation over each time interval. The observed data structure and parameter of interest still write as $O = (W, A, Y^1, Y^2, Y^3, Y^4)$ and $\Psi(P) = (\Psi^j(P))_{1 \le j \le 4}$, but Y^j and $\Psi^j(P)$ now belong to $\mathbb{R}^8$ (and not $\mathbb{R}^3$ anymore). The ranking of the protocols now relies on the criterion $\sum_{i=1}^8 (T_{i,n}^j)^2$, whose definition straightforwardly extends that of the criterion introduced in Section 3.2. The values of the criteria are reported in Table 7. The ranking of protocols remains unchanged, but the discrepancies between the values for protocol 2 on one hand and for protocols 1 and 4 on the other hand are smaller.

protocol	$j = 3$	$j = 2$	$j = 1$	$j = 4$
criterion $\sum_{i=1}^{8}(T_{i,n}^{j})^2$	83.64	43.61	14.92	12.60

Table 7: Ranking the four protocols using the entire real dataset *and the extended small-dimensional summary measure* of the complex trajectories. We report the realizations of the criteria $\sum_{i=1}^{8}(T_{i,n}^{j})^2$ obtained for protocols $j = 1, 2, 3, 4$. The ranking is the same as that derived from Table 5.

We finally apply once again steps 2a, 2b, 2c from the leave-one-out rule described in Section 4.2 where we substitute the real dataset $O_{(1)}, \ldots, O_{(n)}$ for the simulated one, and use either all estimators or only two of them in the super-learner (we exposed our motives in the previous section; see Section A.1 for details on the super-learning procedure). The results are reported in Table 8.

	$J = 1$	$J = 2$	$J = 3$	$J = 4$
Perf^{J} (larger collection)	0.82 (44/54)	0.80 (43/54)	0.80 (43/54)	0.78 (42/54)
Perf^{J} (smaller collection)	0.87 (47/54)	0.85 (46/54)	0.80 (43/54)	0.82 (44/54)

Table 8: Leave-one-out performances Perf^{J} of the classification procedure using the real dataset *and the extended small-dimensional summary measure* of the complex trajectories. Performance Perf^{J} corresponds to the classifier based on J among the four vectors Y^1, Y^2, Y^3, Y^4 (those associated with the J more informative protocols) and either using all estimators (second row) or only two of them (third row) in the super-learner (see Section A.1 for details).

When we include all estimators in the super-learner, the classification procedure that relies on the extended small-dimensional summary measure of the complex trajectories outperforms the classification procedure that relies on the initial summary measure, for every value of J (*i.e.*, each number of protocols taken into account in the classification procedure). The performances are even better when we only include two estimators. Remarkably, the best performance is achieved using only the most informative protocol, with a proportion of subjects correctly classified (evaluated by the leave-one-out rule) equal to 87% (47 out of the 54 subjects are correctly classified)

A   Appendix

We gather in Section A.1 a short and self-contained description of the construction of a super-learner, as well as the estimation procedures that we choose to involve. One of those estimation procedures, a variant of the top-scoring pairs classification procedure, is presented in Section A.2.

A.1   Specifics of the super-learning procedures

We refer to [15, 12] for a general presentation of the super-learning methodology, which is a method for combining/aggregating, by V-fold cross-validation, a family of candidates estimation procedures (or simply *estimators*) of a regression function.

Let us denote by $X \in \mathbb{R}^d$ the vector of predictors and by $Y \in \mathbb{R}$ the outcome of interest (in the classification framework, $Y \in \{0, 1\}$), and let $(X_{(1)}, Y_{(1)}), \ldots, (X_{(n)}, Y_{(n)})$ be n independent copies of (X, Y). The empirical distribution of the whole sample is denoted by P_n. For each $k \in \mathcal{K}$ a finite set of cardinality K, $f_k(P_n)$ is an estimator of the regression function $E(Y|X)$ built on P_n. The objective is to aggregate the K estimators into a single one which will enjoy some optimality/oracle properties with respect to a certain criterion of interest. The criterion of interest drives the choice of a loss function L, which maps a generic observation (X, Y) and a candidate estimator f of the regression function to a real number $L((X, Y), f)$. We use two different loss functions, one for the estimation of the conditional means $Q_i^j(P_0)$ and the conditional distribution $g(P_0)$ as needed to rank the protocols (see Section 3.2), one for the classification of subjects (see Section 3.3). Regarding how to combine $(f_k(P_n))_{k \in \mathcal{K}}$, we decide to consider convex combinations.

We use the value $V = 10$, and therefore draw a n-tuple $(V_{(1)}, \ldots, V_{(n)}) \in \{1, \ldots, V\}^n$ such that $\max_{v, v' \leq V} |\sum_{i=1}^n \mathbf{1}\{V_{(i)} = v\} - \sum_{i=1}^n \mathbf{1}\{V_{(i)} = v'\}| \leq 1$ (the v-strata have the same cardinality, up to one unit), and (in the classification framework) $\min_{v \leq V, y=0,1} \sum_{i=1}^n \mathbf{1}\{Y_{(i)} = y, V_{(i)} = v\} \geq 1$ (each v-stratum contains at least one observation from each class). For every $v \leq V$, let us denote by P_n^v the empirical distribution of those observations for which $V_{(i)} \neq v$; the training set made of the latter observations yields K estimators $(f_k(P_n^v))_{k \in \mathcal{K}}$. Now, we define the minimizer of the V-fold cross-validated risk:

$$\alpha^*(P_n) = \arg\max_{\alpha \in \mathcal{S}_K} \sum_{v \leq V} \sum_{i=1}^n L\left((X_{(i)}, Y_{(i)}), \sum_{k \in \mathcal{K}} \alpha_k f_k(P_n^v)\right) \mathbf{1}\{V_{(i)} = v\}$$

(where $\mathcal{S}_K = \{u \in \mathbb{R}_+^K : \sum_{k \leq K} u_k = 1\}$), which finally yields the *super-learner* obtained as the $\alpha^*(P_n)$-convex combination of the K estimators $(f_k(P_n))_{k \in \mathcal{K}}$ trained on the whole sample, $f^*(P_n) = \sum_{k \in \mathcal{K}} \alpha_k^*(P_n) f_k(P_n)$.

We now turn to the description of the loss function L and of the estimators $(f_k)_{k \in \mathcal{K}}$ specifically used in this article.

- **Estimation of the conditional means $Q_i^j(P_0)$ and the conditional distribution $g(P_0)$.** We choose the squared error loss function, characterized by $L_2((X, Y), f) = (Y - f(X))^2$, whose expectation is minimized, over the set of measurable functions of X, at the targeted regression function $E(Y|X)$.

  Regarding the estimators $(f_k)_{k \in \mathcal{K}}$,

  - **conditional means $Q_i^j(P_0)$:** we rely on (alphabetical order) the elastic net [13, 14], general additive models [8], linear models, loess local regressions [9], random forests [4], and support vector machines [6]. Different values of the various tuning parameters are considered.

  - **conditional distribution $g(P_0)$:** we rely on (alphabetical order) the k-nearest-neighbors, logistic linear regressions, random forests, and support vector machines. Different values of the various tuning parameters are considered.

- **Classification of subjects.** We choose the loss function characterized by $L_3((X, Y), f) = (Y - \text{expit}\{\beta(f(X) - \frac{1}{2})\})^2$ with $\beta = 200$ a large positive number. This loss function is meant to provide a trade-off between the loss functions characterized by $L_1((X, Y), f) = (Y - \mathbf{1}\{f(X) \geq \frac{1}{2}\})^2$ and $L_2((X, Y), f) = (Y - f(X))^2$. Since L_3 is very close to L_1, the super-learner optimizes the convex-combination parameter $\alpha^*(P_n)$ to be applied to the collection $(f_k(P_n))_{k \in \mathcal{K}}$ of estimators in view of the plug-in rule that will ultimately be used (see Section 3.3). Yet L_3 is smoother than L_1 (which takes its values in $\{0, 1\}$), thus in that sense not so far away from L_2, which makes the numerical computation of $\alpha^*(P_n)$ easier for deriving the super-learner.

  Regarding the estimators $(f_k)_{k \in \mathcal{K}}$, we rely on (alphabetical order) the k-nearest-neighbors, logistic regressions, random forests, and the top-scoring pairs classification procedure (see Section A.2). Different values of the tuning parameters are considered for the k-nearest-neighbors and random forests. We also consider the smaller collection of estimators that reduces to random forests (with a single choice of the tuning parameters) and the top-scoring pairs classification procedure.

  Interestingly, the top-scoring pairs classification procedure involves pairwise comparisons $\mathbf{1}\{X^j \leq X^{j'}\}$ $(1 \leq j \neq j' \leq d)$ of predictors: how relevant such comparisons may be for the sake of classification is not taken into account by our method to rank protocols by decreasing order of informativeness relative to postural control.

The R [11] coding of our procedure was eased by the following R-packages: `SuperLearner`, `e1071`, `gam`, `glmnet`, `randomForest` and `stats` (we refer to The Comprehensive R Archive Network, `www.cran.r-project.org/`, for details).

A.2 The top-scoring pairs classification procedure

The top-scoring pair (abbreviated to TSP) classification procedure was introduced in [7] for the purpose of molecular classification based on some genetic information. It is parameter-free and simply relies on pairwise comparisons of predictors, hence its remarkable robustness. Even though the context of the present article greatly differs from that of [7] (mostly in that the number of predictors is small here whereas it is huge there), the very good performances enjoyed by the TSP classifier for molecular classification motivate making a variant of the TSP classification procedure one of the candidates estimators in the super-learner.

Let us denote $X = (X^1, \ldots, X^d) \in \mathbb{R}^d$ the vector of predictors and $Y \in \{0, 1\}$ the class membership indicator. The objective is to estimate the regression function $P(Y = 1|X)$ based on n independent copies $(X_{(1)}, Y_{(1)}), \ldots, (X_{(n)}, Y_{(n)})$ of $(X, Y) \sim P$. One first derives the so-called TSP,

$$(k_0, \ell_0) = \operatorname*{arg\,max}_{1 \leq k < \ell \leq d} \left| \frac{\sum_{i=1}^n \mathbf{1}\{X_{(i)}^k \leq X_{(i)}^\ell, Y_{(i)} = 1\}}{\sum_{i=1}^n \mathbf{1}\{Y_{(i)} = 1\}} - \frac{\sum_{i=1}^n \mathbf{1}\{X_{(i)}^k \leq X_{(i)}^\ell, Y_{(i)} = 0\}}{\sum_{i=1}^n \mathbf{1}\{Y_{(i)} = 0\}} \right|$$

(we implicitly assume that the latter arg max reduces to a single pair; the procedure easily extends to the case that the arg max contains several pairs by making them vote, see [7]). The TSP (k_0, ℓ_0) is chosen in such a way that the empirical conditional probabilities of having $X^{k_0} \leq X^{\ell_0}$ given $Y = 1$ or $Y = 0$ differ as much as possible. Introduce now

$$p_{0,n}^- = \frac{\sum_{i=1}^n \mathbf{1}\{X_{(i)}^{k_0} \leq X_{(i)}^{\ell_0}, Y_{(i)} = 1\}}{\sum_{i=1}^n \mathbf{1}\{X_{(i)}^{k_0} \leq X_{(i)}^{\ell_0}\}},$$

$$p_{0,n}^+ = \frac{\sum_{i=1}^n \mathbf{1}\{X_{(i)}^{k_0} > X_{(i)}^{\ell_0}, Y_{(i)} = 1\}}{\sum_{i=1}^n \mathbf{1}\{X_{(i)}^{k_0} > X_{(i)}^{\ell_0}\}}.$$

Our TSP estimator of the regression function $P(Y = 1|X)$ finally writes as

$$P_n^{\mathrm{TSP}}(Y = 1|X) = p_{0,n}^- \mathbf{1}\{X^{k_0} \leq X^{\ell_0}\} + p_{0,n}^+ \mathbf{1}\{X^{k_0} > X^{\ell_0}\}.$$

An alternative definition (closer to the original TSP procedure from [7]) could have been to introduce $\pi_{0,n}^y = \frac{\sum_{i=1}^n \mathbf{1}\{X_{(i)}^{k_0} \leq X_{(i)}^{\ell_0}, Y_{(i)} = y\}}{\sum_{i=1}^n \mathbf{1}\{Y_{(i)} = y\}}$ for $y = 0, 1$ and to estimate the regression function $P(Y = 1|X)$ by $\pi_{0,n}^1 \mathbf{1}\{\pi_{0,n}^1 \geq \pi_{0,n}^0\} + (1 - \pi_{0,n}^0)\mathbf{1}\{\pi_{0,n}^1 < \pi_{0,n}^0\}$ if $X^{k_0} \leq X^{\ell_0}$, and by $(1 - \pi_{0,n}^0)\mathbf{1}\{\pi_{0,n}^1 \geq \pi_{0,n}^0\} + \pi_{0,n}^1 \mathbf{1}\{\pi_{0,n}^1 < \pi_{0,n}^0\}$ otherwise.

References

[1] J-Y. Audibert and A. B. Tsybakov. Fast learning rates for plug-in classifiers. *Ann. Statist.*, 35(2):608–633, 2007.

[2] P. Bertrand, J-M. Bardet, M. Dabonneville, A. Mouzat, and P. Vaslin. Automatic determination of the different control mechanisms in upright position by a wavelet method. *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, 2:1163 – 1166, 2001.

[3] I. Bonan, A. Yelnik, I. Laffont, E. Vitte, and G. Freyss. Selection of sensory information in postural control of hemiplegics after unique stroke. *Annales de Réadaptation et de Médecine Physique*, 39, Issue 3:157–163, 1996.

[4] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[5] A. Chambaz, I. Bonan, and P-P. Vidal. Deux modèles de Markov caché pour processus multiples et leur contribution à l'élaboration d'une notion de style postural. *J. SFdS*, 150(1):73–100, 2009.

[6] N. Christianini and J. Shawe-Taylor. *An introduction to Support Vector Machines*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000.

[7] D. Geman, C. d'Avignon, D. Q. Naiman, and R. L. Winslow. Classifying gene expression profiles from pairwise mRNA comparisons. *Stat. Appl. Genet. Mol. Biol.*, 3, 2004. Article 19.

[8] T. J. Hastie and R. J. Tibshirani. *Generalized additive models*, volume 43 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London, 1990.

[9] C. Loader. *Local Regression and Likelihood*. Statistics and Computing. Springer, New-York, 1999.

[10] K. M. Newell, S. M. Slobounov, E. S. Slobounova, and P. C. M. Molenaar. Stochastic processes in postural center-of-pressure profiles. *Experimental Brain Research*, 113:158–164, 1997.

[11] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. ISBN 3-900051-07-0.

[12] S. Rose and M. J. van der Laan, editors. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer Verlag, 2011.

[13] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996.

[14] R. Tibshirani, T. Hastie, and J. H. Friedman. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.*, 33, Issue 1, 2010.

[15] M. J. van der Laan, E. C. Polley, and A. E. Hubbard. Super learner. *Stat. Appl. Genet. Mol. Biol.*, 6:Art. 25, 23 pp. (electronic), 2007.

[16] M. J. van der Laan and D. Rubin. Targeted maximum likelihood learning. *Int. J. Biostat.*, 2:Art. 11, 40, 2006.