Nonlinear functional regression: a functional RKHS approach

Hachem Kadri, Emmanuel Duflos, Philippe Preux, Stéphane Canu and Manuel Davy

INRIA Lille - Nord Europe

May 13, 2010

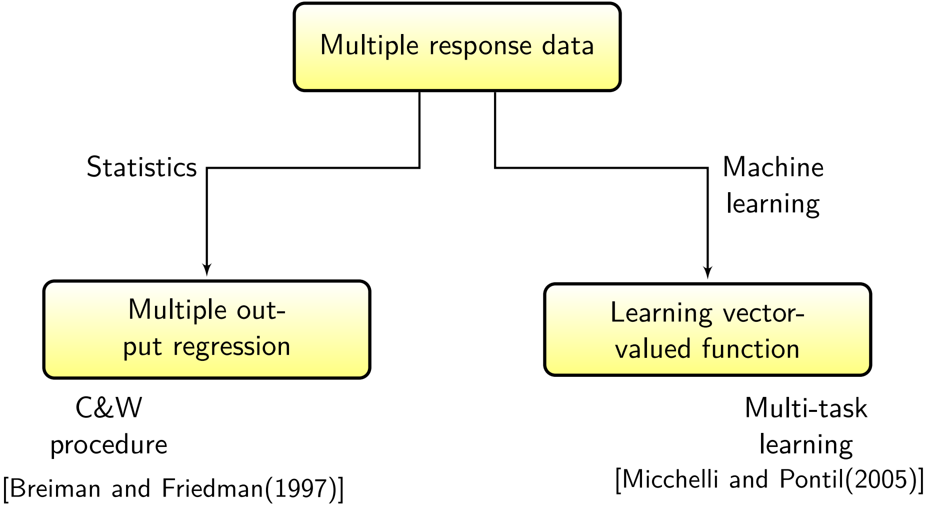
Outline

Regression

$$y_i = f(x_i) + \epsilon_i$$

Predictor $\longmapsto$	Response	Model
$\mathbb{R}^d$	$\mathbb{R}$	Multiple Regression, GLM
$\mathbb{R}^d$	$\mathbb{R}^d$	Multivariate Regression
$\mathbf{L}^2$	$\mathbb{R}$	Generalized Functional Linear Model
$\mathbb{R}^d$	$\mathbf{L}^2$	Functional Response Model
$\mathbf{L}^2$	$\mathbf{L}^2$	Functional Regression Model

Multiple response data



Functional Data Analysis (FDA) - FAQ -

- **What is functional data analysis ? goals ?**
  The study of curves, shapes or other high-dimensional objects. [Ramsay and Silverman(2002), Ramsay and Silverman(2005)]
- **What is functional data ?**
  Represent a number of discrete observations about an individual in the form of a function or curve.
- **Where to find such data ?**
  Meteorology / Finance / Biology / Chemometrics / Speech analyis / Brain imaging
- **How to treat these data ?**
  3 schools: English - French - Stochastic
- **Multivariate data or Functional data ?**
  No answer ??? !!!

Functional Data Analysis (FDA) - FAQ -

- **What is functional data analysis ? goals ?**

  The study of curves, shapes or other high-dimensional objects. [Ramsay and Silverman(2002), Ramsay and Silverman(2005)]

- **What is functional data ?**

  Represent a number of discrete observations about an individual in the form of a function or curve.

- **Where to find such data ?**

  Meteorology / Finance / Biology / Chemometrics / Speech analyis / Brain imaging

- **How to treat these data ?**

  3 schools: English - French - Stochastic

- **Multivariate data or Functional data ?**

  No answer ??? !!!

Functional Data Analysis (FDA) - FAQ -

- **What is functional data analysis ? goals ?**
  The study of curves, shapes or other high-dimensional objects. [Ramsay and Silverman(2002), Ramsay and Silverman(2005)]
- **What is functional data ?**
  Represent a number of discrete observations about an individual in the form of a function or curve.
- **Where to find such data ?**
  Meteorology / Finance / Biology / Chemometrics / Speech analyis / Brain imaging
- **How to treat these data ?**
  3 schools: English - French - Stochastic
- **Multivariate data or Functional data ?**
  No answer ??? !!!

Functional Data Analysis (FDA) - FAQ -

- **What is functional data analysis ? goals ?**

  The study of curves, shapes or other high-dimensional objects. [Ramsay and Silverman(2002), Ramsay and Silverman(2005)]

- **What is functional data ?**

  Represent a number of discrete observations about an individual in the form of a function or curve.

- **Where to find such data ?**

  Meteorology / Finance / Biology / Chemometrics / Speech analyis / Brain imaging

- **How to treat these data ?**

  3 schools: English - French - Stochastic

- **Multivariate data or Functional data ?**

  No answer ??? !!!

Functional Data Analysis (FDA) - FAQ -

- **What is functional data analysis ? goals ?**
  The study of curves, shapes or other high-dimensional objects. [Ramsay and Silverman(2002), Ramsay and Silverman(2005)]
- **What is functional data ?**
  Represent a number of discrete observations about an individual in the form of a function or curve.
- **Where to find such data ?**
  Meteorology / Finance / Biology / Chemometrics / Speech analyis / Brain imaging
- **How to treat these data ?**
  3 schools: English - French - Stochastic
- **Multivariate data or Functional data ?**
  No answer ??? !!!

Summary of Motivations

- Nonlinear functional methods
- High dimensional data
- Correlation between data
- Time information - sequential tools -

Functional kernel

$(x_i(s), y_i(t))_{i=1}^n \in \mathcal{G}_x \times \mathcal{G}_y$ where $\mathcal{G}_x : \Omega_x \longrightarrow \mathbb{R}$ and $\mathcal{G}_y : \Omega_y \longrightarrow \mathbb{R}$

Definition

An $\mathcal{L}(\mathcal{G}_y)$-valued kernel $K_\mathcal{F}(w, z)$ on $\mathcal{G}_x$ is a function $K_\mathcal{F}(.,.) : \mathcal{G}_x \times \mathcal{G}_x \longrightarrow \mathcal{L}(\mathcal{G}_y)$;

- *$K_\mathcal{F}$ is Hermitian if $K_\mathcal{F}(w, z) = K_\mathcal{F}(z, w)^*$,*
- *it is nonnegative on $\mathcal{G}_x$ if for any $\{(w_i, u_i)_{i=1,\ldots,r}\} \in \mathcal{G}_x \times \mathcal{G}_y$*

$$\sum_{i,j} \langle K_\mathcal{F}(w_i, w_j) u_i, u_j \rangle_{\mathcal{G}_y} \geq 0$$

Functional RKHS (1/4)

Definition

A Hilbert space $\mathcal{F}$ of functions from $\mathcal{G}_x$ to $\mathcal{G}_y$ is called a reproducing kernel Hilbert space if there is a nonnegative $\mathcal{L}(\mathcal{G}_y)$-valued kernel $K_{\mathcal{F}}(w, z)$ on $\mathcal{G}_x$ such that:

- *the function $z \longmapsto K_{\mathcal{F}}(w, z)g$ belongs to $\mathcal{F}$ for every choice of $w \in \mathcal{G}_x$ and $g \in \mathcal{G}_y$*
- *for every $f \in \mathcal{F}, \langle f, K_{\mathcal{F}}(w, .)g\rangle_{\mathcal{F}} = \langle f(w), g\rangle_{\mathcal{G}_y}$*

(reproducing property)

Functional RKHS (2/4)

Theorem

If a Hilbert space $\mathcal{F}$ of functions on $\mathcal{G}_y$ admits a reproducing kernel, then the reproducing kernel $K_{\mathcal{F}}(w,z)$ is uniquely determined by the Hilbert space $\mathcal{F}$.

Proof:

$$\langle K'(w',.)h, K(w,.)g\rangle_{\mathcal{F}} = \langle K'(w',w)h, g\rangle_{\mathcal{G}} \tag{1}$$

$$\begin{aligned} \langle K'(w',.)h, K(w,.)g\rangle_{\mathcal{F}} &= \langle K(w,.)g, K'(w',.)h\rangle_{\mathcal{F}} \\ &= \langle K(w,w')g, h\rangle_{\mathcal{G}} \\ &= \langle g, K(w,w')^{*}h\rangle_{\mathcal{G}} \\ &= \langle g, K(w',w)h\rangle_{\mathcal{G}} \end{aligned} \tag{2}$$

(1) and $(2) \Longrightarrow K_{\mathcal{F}}(w,z) \equiv K'_{\mathcal{F}}(w,z)$ □

Functional RKHS (3/4)

Theorem

A $\mathcal{L}(\mathcal{G}_y)$-valued kernel $K_{\mathcal{F}}(w,z)$ on $\mathcal{G}_x$ is the reproducing kernel of some Hilbert space $\mathcal{F}$, if and only if it is positive definite.

Proof: Function-valued RKHS ⇒ positive operator-valued kernel

$$\sum_{i,j=1}^{n} \langle K_{\mathcal{F}}(w_i,w_j),u_j\rangle_{\mathcal{G}_y}$$
$$= \sum_{i,j=1}^{n} \langle K_{\mathcal{F}}(w_i,.)u_i, K_{\mathcal{F}}(w_j,.)u_j\rangle_{\mathcal{F}}$$
$$= \langle \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,.)u_i, \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,.)u_i\rangle_{\mathcal{F}}$$
$$= \| \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,.)u_i\|^2_{\mathcal{F}} \geq 0 \qquad \square$$

Functional RKHS (4/4)

Proof: Function-valued RKHS $\Leftarrow$ positive operator-valued kernel

- $\mathcal{F}_0, \ \forall f \in \mathcal{F}_0, \ f(.) = \sum_{i=1}^{n} K_{\mathcal{F}}(w_i, .)\alpha_i$
- $\langle f(.), g(.)\rangle_{\mathcal{F}_0} = \sum_{i,j=1}^{n} \langle K_{\mathcal{F}}(w_i, z_j)\alpha_i, \beta_j\rangle_{\mathcal{G}_y}$
- $(\mathcal{F}_0, \langle ., .\rangle_{\mathcal{F}_0})$ is a pre-Hilbert space
- $\mathcal{F}$ the completion of $\mathcal{F}_0$ space via Cauchy sequences

Functional kernel construction

- $\mathcal{L}(\mathcal{G})$-valued kernel

$$\begin{array}{rccc} K_{\mathcal{F}}: & \mathcal{G}\times\mathcal{G} & \longrightarrow & \mathcal{L}(\mathcal{G}) \\ & x,y & \longmapsto & T^{h(x,y)} \end{array}$$

- Operator

$$\begin{array}{rccc} T^{h}: & \mathcal{G} & \longrightarrow & \mathcal{G} \\ & x & \longmapsto & T_x^h \quad ; \quad T_x^h(t) = h(t)x(t) \end{array}$$

- characterestic function

$$\begin{array}{rccc} K_{\mathcal{F}}: & \mathcal{G}\times\mathcal{G} & \longrightarrow & \mathcal{G} \\ & x,y & \longmapsto & h(x,y) = \exp(c.(x-y)^2) \end{array}$$

Representer theorem (1/2)

Theorem

The solution of the minimization problem

$$\min_{f \in \mathcal{F}} \sum_{i=1}^{n} \|y_i - f(x_i)\|^2_{\mathcal{G}_y} + \lambda \|f\|^2_{\mathcal{F}}$$

is achieved by a function of the form

$$f^*(.) = \sum_{i=1}^{n} K_{\mathcal{F}}(x_i, .)\beta_i$$

Representer theorem (2/2)

Proof:

$$D_h J_\lambda(f) = \lim_{\tau \longrightarrow 0} \frac{J_\lambda(f+\tau h) - J_\lambda(f)}{\tau} = \langle J_\lambda^{'}(f), h \rangle$$

i. $G(f) = \|f\|_{\mathcal{F}}^2$

$$\lim_{\tau \longrightarrow 0} \frac{\|f+\tau h\|_{\mathcal{F}}^2 - \|f\|_{\mathcal{F}}^2}{\tau} = 2\langle f, h \rangle \qquad \Longrightarrow \qquad G^{'}(f) = 2f$$

ii. $H_i(f) = \|y_i - f(x_i)\|_{\mathcal{G}}^2$

$$\lim_{\tau \longrightarrow 0} \frac{\|y_i - f(x_i) - \tau h(x_i)\|_{\mathcal{G}}^2 - \|y_i - f(x_i\|_{\mathcal{G}}^2}{\tau}$$
$$= -2\langle y_i - f(x_i), h(x_i)\rangle_{\mathcal{G}} = -2\langle K_{\mathcal{F}}(x_i,.)(y_i - f(x_i)), h\rangle_{\mathcal{F}}$$
$$= -2\langle K_{\mathcal{F}}(x_i,.)\alpha_i, h\rangle_{\mathcal{F}} \qquad \Longrightarrow \qquad H_i^{'}(f) = 2K_{\mathcal{F}}(x_i,.)\alpha_i$$

(i), (ii), and $J_\lambda^{'}(f^*) = 0 \Longrightarrow f^*(.) = \frac{1}{\lambda}\sum_{i=1}^{n} K_{\mathcal{F}}(x_i,.)\alpha_i$ □

Regression function estimation (1/2)

$$\min_{f \in \mathcal{F}} \sum_{i=1}^{n} \|y_i - f(x_i)\|_{\mathcal{G}_y}^2 + \lambda \|f\|_{\mathcal{F}}^2$$

using the representer theorem

$$\Longleftrightarrow \quad \min_{\beta_i} \sum_{i=1}^{n} \|y_i - \sum_{j=1}^{n} K_{\mathcal{F}}(x_i, x_j)\beta_j\|_{\mathcal{G}_y}^2 + \lambda \|\sum_{j=1}^{n} K_{\mathcal{F}}(., x_j)\beta_j\|_{\mathcal{F}}^2$$

using the reproducing property

$$\Longleftrightarrow \quad \min_{\beta_i} \sum_{i=1}^{n} \|y_i - \sum_{j=1}^{n} K_{\mathcal{F}}(x_i, x_j)\beta_j\|_{\mathcal{G}_y}^2 + \lambda \sum_{i,j}^{n} \langle K_{\mathcal{F}}(x_i, x_j)\beta_i, \beta_j \rangle_{\mathcal{G}_y}$$

Regression function estimation (2/2)

$$\min_{\beta_i} \sum_{i=1}^{n} \|y_i - \sum_{j=1}^{n} K_{\mathcal{F}}(x_i, x_j)\beta_j\|^2_{\mathcal{G}_y} + \lambda \sum_{i,j}^{n} \langle K_{\mathcal{F}}(x_i, x_j)\beta_i, \beta_j\rangle_{\mathcal{G}_y}$$

Multiplication operator

$$\iff \quad \min_{\beta_i} \sum_{i=1}^{n} \|y_i - \sum_{j=1}^{n} c_{ij}\beta_j\|^2_{\mathcal{G}_y} + \lambda \sum_{i,j}^{n} \langle c_{ij}\beta_i, \beta_j\rangle_{\mathcal{G}}$$

Discrete points

$$\iff \quad \min_{\beta_i} \sum_{i=1}^{n} \sum_{l=1}^{p} \left(y_i(t_i^l) - \sum_{j=1}^{n} c_{ij}(t_{ij}^l)\beta_j(t_j^l) \right)^2 + \lambda \sum_{i,j}^{n} \sum_{l}^{p} c_{ij}(t_{ij}^l)\beta_i(t_i^l)\beta_j(t_j^l)$$

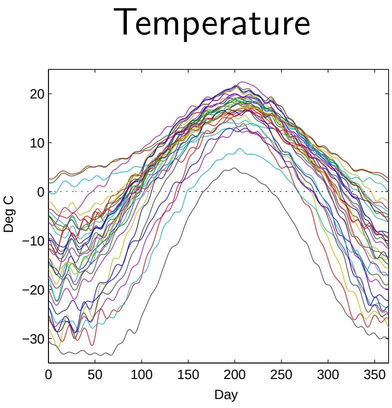
Matrix formulation

$$\iff \quad \min_{\beta} trace((Y - C\beta)(Y - C\beta)^T) + \lambda\, trace(C\beta\beta^T)$$

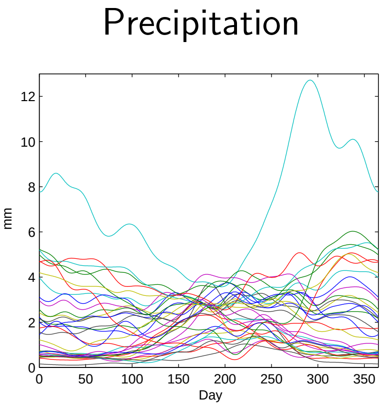
Matrix derivative

$$\implies \quad (C + \lambda I)\beta = Y$$

Meteorological data

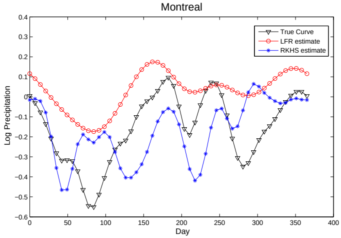


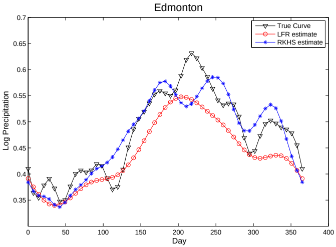
(a)

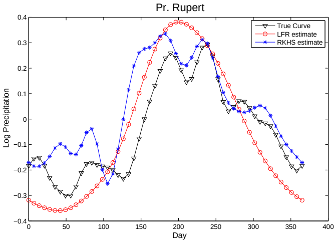


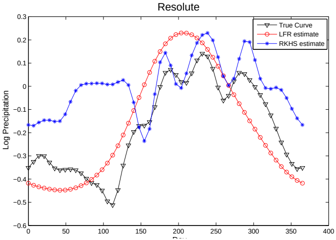
(b)

Prediction results









Conclusion

- Nonlinear functional regression
- RKHS - operator sapces
- Multiple functional regression
- Operator-valued kernels
- Regularization path algorithm

L. Breiman and J. Friedman.
Predicting multivariate responses in multiple linear regression.
Journal of the Royal Statistical Society, Series B, 59:3–54, 1997.

C. A. Micchelli and M. Pontil.
Kernels for multi-task learning.
In *Advances in Neural Information Processing Systems 17 (NIPS 2005)*, pages 921–928, 2005.

J. O. Ramsay and B. W. Silverman.
Applied Functional Data Analysis.
Springer Verlag, New York, 2002.

J. O. Ramsay and B. W. Silverman.
Functional Data Analysis, 2nd edn.
Springer Verlag, New York, 2005.