RKHSMetaMod : An R package to estimate the Hoeffding decomposition of an unknown function by solving RKHS Ridge Group Sparse optimization problem

Halaleh KAMARI

halaleh.kamari@univ-evry.fr

Résumé. RKHSMetaMod est un package R qui estime un méta-modèle d'une fonction inconnue m dans le cadre d'un modèle de régression Gaussien. La procédure repose sur la minimisation d'un critère des moindres carrés pénalisé par une double pénalité dite Ridge Group Sparse, pour des fonctions appartenant à un espace de Hilbert a Noyau Reproduisant (RKHS). Le méta-modèle estimé est un modéle additif dont les termes estiment les termes de la décomposition de Hoeffding de la fonction m. Ce package fournit une interface conviviale entre l'environnement informatique statistique R et les bibliothèques C++ Eigen et GSL. Le temps d'exécution est optimisé via l'utilisation des packages RcppEigen et RcppGSL.
Mots-clés. méta-modèle, décomposition de Hoeffding, pénalité Ridge Group Sparse, espaces de Hilbert a Noyau Reproduisant.

Abstract. RKHSMetaMod is an R package that fits a meta model to an unknown model m by solving the Ridge Group Sparse Optimization Problem based on Reproducing Kernel Hilbert Spaces (RKHS). The obtained meta model is an additive model that satisfies the properties of the Hoeffding decomposition, and its terms estimate the terms in the Hoeffding decomposition of the function m. This package provides an interface from R statistical computing environment to the C++ libraries Eigen and GSL. It uses the performance C++ functions through RcppEigen and RcppGSL packages to speed up the execution time and the R environment in order to propose an user friendly package.
Keywords. meta model, Hoeffding decomposition, Ridge Group Sparse penalty, Reproducing Kernel Hilbert Spaces.

1 Introduction

Consider the Gaussian regression model $Y = m(X) + \sigma\varepsilon$, $\sigma \in \mathbb{R}$. The variables $X_1, ..., X_d$ are independently distributed with law P_X on $\mathcal{X}$ and are independent of ε's, the function m is unknown and square integrable, i.e. $m \in L^2(P_X, \mathcal{X})$.
Since the inputs X are independent the function $m(X)$ can be written according to its Hoeffding decomposition, see Sobolá (2001). Let $\mathcal{P}$ be the set of parts of $\{1, ..., d\}$

with dimension 1 to d, and X_v the set of variables for all $v \in \mathcal{P}$, then the Hoeffding decomposition of m is written as :

$$m(X) = m_0 + \sum_{v \in \mathcal{P}} m_v(X_v), \tag{1}$$

where m_0 is a constant, and m_v is a function of X_v. In this decomposition all the terms are orthogonal with respect to $L^2(P_X, \mathcal{X})$.
Thanks to the independency between the variables X_a for $a = 1, ..., d$, the variance of $m(X)$ can be decomposed as follows :

$$\mathrm{Var}(m(X)) = \sum_v \mathrm{Var}(m_v).$$

The sensitivity indices, introduced by Sobolá (2001), are defined for any group of variables $X_v, v \in \mathcal{P}$ by :

$$S_v = \frac{\mathrm{Var}(m_v(X_v))}{\mathrm{Var}(m(X))}.$$

Since the function m is unknown, the functions m_v are also unknown. The idea is to approximate the Hoeffding decomposition of m by an additive model, called RKHS meta model. This meta model belongs to the RKHS $\mathcal{H} = \bigotimes \mathcal{H}_v$, constructed as proposed by Durrande et al. (2013).
Thanks to the properties of this space any function $f \in \mathcal{H}$, satisfies :

$$f(X) =< f, k(X, .) >_{\mathcal{H}} = f_0 + \sum_v f_v(X),$$

where $< ., . >_{\mathcal{H}}$ denotes the scalar product in $\mathcal{H}$, k is the reproducing kernel associated with the RKHS $\mathcal{H}$, and $f_v(X) =< f, k_v(X, .) >_{\mathcal{H}}$ for k_v being the reproducing kernel associated with the RKHS $\mathcal{H}_v$.
Moreover, for all $v \in \mathcal{P}$, $f_v(X_v)$ are centered and for all $v \neq v'$ the functions $f_v(X_v)$ and $f_{v'}(X_{v'})$ are uncorrelated. So, we get the Hoeffding decomposition of f.
The function m is approximated by the RKHS meta model $f^* \in \mathcal{H}$, which is the solution of the residual sum of squares minimization, penalized by a Ridge Group Sparse penalty function. This method both estimates the groups v that are suitable for predicting m and the relationship between m_v and X_v. The terms of the Hoeffding decomposition of the estimator f^* are the approximations of each $m_v(X_v)$ in Equation (1).
RKHSMetaMod is an R package, that implements the RKHS Ridge Group Sparse optimization algorithm to approximate the terms m_v in the decomposition given at Equation (1). As a consequence we get an estimation of the function m.
To be more precise, it provides the functions RKHSgrplasso() and RKHSMetMod() to fit a solution of the two following convex optimization problems :

- RKHS Group Lasso,
- RKHS Ridge Group Sparse,

where RKHS Group Lasso is a special case of RKHS Ridge Group Sparse algorithm. These algorithms are described in the next section. In section 3 we give an overview of the main function of the package and we illustrate the use of this function through an example.

2 Description of the method

2.1 RKHS Ridge Group Sparse Optimization Problem

Let n be the number of observations. The dataset consists of Y, vector of n observations, and X, an $n \times d$ matrix of features, with components $(X_{ai}, i = 1, ..., n, a = 1, ..., d) \in \mathbb{R}^{n\times d}$. For some tuning parameters γ_v and μ_v, we consider the RKHS Ridge Group Sparse criteria,

$$\frac{1}{n}\|Y - f_0 - \sum_{v\in\mathcal{P}} f_v(X_v)\|^2 + \frac{1}{\sqrt{n}}\gamma \sum_{v\in\mathcal{P}} \|f_v\| + \mu \sum_{v\in\mathcal{P}} \|f_v\|_{\mathcal{H}_v}, \tag{2}$$

where $\|.\|$ is the Euclidean norm in $\mathbb{R}^n$, and matrix X_v represents the predictors corresponding to the v-th group. The minimization of Equation (2) is carried out over a proper subset of $\mathcal{H}$.

According to the Representer Theorem, see Kimeldorf and Wahba (1970), for all $v \in \mathcal{P}$, we have $f_v(.) = \sum_{i=1}^{n} \theta_{vi} k_v(X_{vi,.})$ for some matrix $\theta = (\theta_{vi}, i = 1, ..., n, v \in \mathcal{P}) \in \mathbb{R}^{n\times|\mathcal{P}|}$. Therefore, the minimization of Equation (2) over a set of functions of $\mathcal{H}$ comes to the minimization of Equation (3) over $f_0 \in \mathbb{R}$, and $\theta_v \in \mathbb{R}^n$:

$$C(f_0, \theta_v) = \|Y - f_0 I_n - \sum_{v\in\mathcal{P}} K_v\theta_v\|^2 + \sqrt{n}\gamma \sum_{v\in\mathcal{P}} \|K_v\theta_v\| + n\mu \sum_{v\in\mathcal{P}} \|K_v^{1/2}\theta_v\|, \tag{3}$$

where K_v is the $n \times n$ Gram matrix associated with the kernel $k_v(X_v, .)$, see Huet and Taupin (2017).

2.2 RKHS Group Lasso Optimization Problem

The minimization of Equation (3) could be seen as a Group Lasso optimization problem by considering only the second penalty function, i.e. setting $\gamma = 0$.

The RKHS Group Lasso criteria is defned as :

$$C_g(f_0, \theta_v) = \|Y - f_0 I_n - \sum_{v\in\mathcal{P}} K_v\theta_v\|^2 + n\mu \sum_{v\in\mathcal{P}} \|K_v^{1/2}\theta_v\|. \tag{4}$$

For more details about the ordinary Group Lasso algorithm, see Meier et al. (2008). From now on we denote the penalty parameter in the RKHS Group Lasso algorithm by $\mu_g = \sqrt{n}\mu$.

2.3 Choice of the tuning parameters

We propose to use a sequence of tuning parameters to create a series of estimators. These estimators are evaluated using a testing dataset $(Y_i^{\text{test}}, X_i^{\text{test}}), i = 1, ..., n^{\text{test}}$. For each value of (μ, γ) in the sequence, let $\hat{f}_{(\mu,\gamma)}$ be the estimation of m, obtained by the learning dataset. Then, the prediction error is calculated by,

$$\text{ErrPred}(\mu, \gamma) = \frac{1}{n^{\text{test}}} \sum_{i=1}^{n^{\text{test}}} (Y_i^{\text{test}} - \hat{f}_{(\mu,\gamma)}(X_i^{\text{test}}))^2,$$

where $\hat{f}_{(\mu,\gamma)}(X^{\text{test}}) = \hat{f}_0 + \sum_v \sum_{i=1}^n k_v(X_{vi}, X_v^{\text{test}})\theta_{vi}$.
We choose the pair $(\hat{\mu}, \hat{\gamma})$ with the smallest prediction error, and the model associated with these chosen tuning parameters is the "best" estimator, $\hat{f} = \hat{f}_{(\hat{\mu},\hat{\gamma})}$ of the true model m.
To set up the grid of values of μ, one can set $\gamma = 0$, and find $\mu_{\max}$, the smallest value of μ_g such that the solution to the minimization of the RKHS Group Lasso problem is $\theta_v = 0, \quad \forall v \in \mathcal{P}$. Then $\mu_l = \frac{\mu_{\max}}{\sqrt{n}} \times 2^{-l}, l \in \{1, ..., l_{\max}\}$ could be a grid of values for μ.

2.4 Sensitivity indices

Once we obtain the estimator $\hat{f}$, we calculate it's sensitivity indices by,

$$\hat{S}_v = \frac{\text{Var}(\hat{f}_v(X_v))}{\text{Var}(\hat{f}(X))}.$$

Since $\hat{f} \in \mathcal{H}$, we have $\text{Var}(\hat{f}(X)) = \sum_v \text{Var}(\hat{f}_v(X_v))$. We use an estimator based on the empirical variances of functions $\hat{f}_v$ (Huet and Taupin (2017)) :

$$\widehat{\text{Var}}(\hat{f}_v) = \frac{1}{n-1} \sum_i (\hat{f}_v(X_{v,i}) - \hat{f}_{v,.})^2,$$

where $\hat{f}_{v,.}$ is the mean of $\hat{f}_v(X_{v,i})$ for $i = 1, ..., n$. The $\hat{S}_v$ are the approximations of the sensitivity indices S_v, for all $v \in \mathcal{P}$, of the function m.

3 Main function of the RKHSMetaMod

RKHSMetMod() function Calculates a sequence of meta models which are the solutions of the RKHS Ridge Group Sparse or RKHS Group Lasso optimization problems. In Table 1 the reader will find a summary of all the input parameters of the RKHSMetMod() function and default values for non mandatory parameters.

The RKHSMetMod() function returns an instance of the "RKHSMetMod" class. Its three attributes will contain all outputs :

Input parameter	Description
Y	Vector of response observations of size n.
X	Matrix of input observations with n rows and d columns. Rows correspond to observations and columns correspond to variables.
kernel	Character, indicates the type of reproducing kernel choosed to construct the RKHS $\mathcal{H}$.
Dmax	Integer, between 1 and d, indicates the order of interactions considered in the meta model, i.e. Dmax$=1$ is used to consider only the main effects, Dmax$=2$ to include the main effects and the interactions of order 2,....
gamma	Vector of non negative scalars, values of the penalty parameter γ in decreasing order. If $\gamma = 0$ the function solves the RKHS Group Lasso problem and for $\gamma > 0$ it solves the RKHS Ridge Group Sparse problem.
frc	Vector of positive scalars. Each element of the vector sets a value to the penalty parameter μ, $\mu = \frac{\mu_{\max}}{\sqrt{n} \times \mathrm{frc}}$. The value $\mu_{\max}$ is calculated inside the program.
verbose	Logical, set as FALSE by default.

Table 1: List of the input parameters of RKHSMetMod() function

- mu : Value of the penalty parameter μ or μ_g, depending on the value of the penalty parameter γ.
- gamma : Value of the penalty parameter γ.
- Meta-Model : A RKHS Ridge Group Sparse or RKHS Group Lasso object associated with the penalty parameters mu and gamma.

Ilustration of use of this function is given in the following Example.

Example Simulate the experiment proposed by Durrande et al. (2013) :
Recall our model, $Y = m(X) + \sigma\varepsilon$. We set $\sigma = 0.2$, $\varepsilon \sim \mathcal{N}(0, 1)$, and we consider the g-function of Sobol, see Saltelli et al. (2009), defined over $[0, 1]^d$ by,

$$m(X) = \prod_{a=1}^{d} \frac{|4x_a - 2| + c_a}{1 + c_a}, c_a > 0.$$

Set $n = 100$, $d = 5$, $Dmax = 3$ and $(c_1, c_2, c_3, c_4, c_5) = (0.2, 0.6, 0.8, 100, 100)$. The prediction error for a series of RKHS meta models obtained by the function RKHSMetMod() is displayed in Table 2.

The minimum value of prediction error is obtained for $(\hat{\mu}, \hat{\gamma}) = (0.002, 0.01)$, and the

μ	0.0434	0.0217	0.0108	0.0054	0.0027	0.0013	0.0006
γ_1 =0.2	0.2744	0.1928	0.1608	0.1459	0.1374	0.1383	0.1612
γ_2 =0.1	0.2187	0.1553	0.1329	0.1193	0.0978	0.0913	0.1057
γ_3 =0.01	0.1789	0.1322	0.1188	0.1008	0.08301	0.0831	0.0867
γ_4 =0.005	0.1771	0.1312	0.1183	0.1013	0.08556	0.0892	0.0946
γ_5 =0	0.1751	0.1302	0.1181	0.1021	0.0880	0.0954	0.1065

Table 2: Prediction error.

"best" RKHS meta model is then $\hat{f}_{(0.002,0.01)}$.
The obtained SIs are presented in the Table 3. In the first row the reader finds the true SIs, in the second row the results obtained by Durrande et al. (2011), in the third row the empirical SIs for $\hat{f}_{(0.002,0.01)}$, and the last row is the mean of the empirical SIs of the "best" RKHS meta models for 10 generated experimental designs.

v	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$	sum
SI	0.43	0.24	0.19	0.06	0.04	0.03	0.01	1
SId	0.44	0.24	0.19	0.01	0.01	0.01	0.00	0.9
SI.minErr	0.44	0.27	0.25	0.02	0.01	0.01	0.00	1
mean.SI.minErr	0.46	0.25	0.18	0.04	0.03	0.01	0.00	0.97

Table 3: Sensitivity Indices.

Bibliographie

Durrande, N., Ginsbourger, D., Roustant, O., and Carraro, L. (2013). Anova kernels and rkhs of zero mean functions for model-based sensitivity analysis. Journal of Multivariate Analysis 115, 57 – 67.

Huet, S. and Taupin, M.-L. (2017). Metamodel construction for sensitivity analysis. ESAIM: Procs 60, 27–69.

Kimeldorf, G. S. and Wahba, G. (1970, 04). A correspondence between bayesian estimation on stochastic processes and smoothing by splines. Ann. Math. Statist. 41(2), 495–502.

Meier, L., van de Geer, S., and Buhlmann, P. (2008). The group lasso for logistic regression. Journal of the Royal Statistical Society. Series B 70(1), 53–71.

Saltelli, A., Chan, K., and Scott, E. (2009). Sensitivity Analysis. Wiley.

Sobolá, I. M. (2001, February). Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Math. Comput. Simul. 55(1-3), 271–280.