Optimal Decision Trees for Complex Data

Rafael Blanquero[1] Emilio Carrizosa[1] **Cristina Molero-Río**[1]
Dolores Romero Morales[2]

[1]Instituto de Matemáticas de la Universidad de Sevilla, Seville, Spain
[2]Copenhagen Business School, Frederiksberg, Denmark

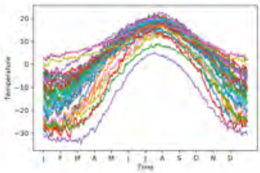
New Bridges between Mathematics and Data Science

November 9th 2021

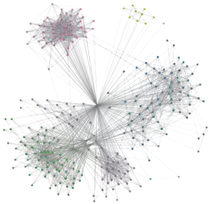
Vectorial

	name	surname	age	weight	height
0	Jorge	Perez	24	50	170
1	Pepe	Garcia	27	60	175
2	Aria	Jimenez	26	70	180
3	Maria	Ruz	25	75	181
4	Luisa	Perez	24	50	170
5	Luisa	Perez	24	50	170

Functional



Network



Image

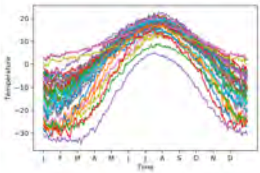
Text


ham What you doing?how are you?
ham Ok lar... Joking wif u oni...
ham dun say so early hor... U c already then say...
ham MY NO. IN LUTON 0125698789 RING ME IF UR AROUND! H*
ham Siva is in hostel aha:-.
ham Cos i was out shopping wif darren jus now n i called him 2 ask wat present
wan lor. Then he started guessing who i was wif n he finally guessed darren lor.
spam FreeMsg: Txt: CALL to No: 86888 & claim your reward of 3 hours talk time to
use from your phone now! ubscribe6GBP/ mnth inc 3hrs 16 stop?txtStop
spam Sunshine Quiz! Win a super Sony DVD recorder if you canname the capital of
Australia? Text MQUIZ to 82277. B
spam URGENT! Your Mobile No 07808726822 was awarded a L2,000 Bonus Caller Prize
02/09/03! This is our 2nd attempt to contact YOU! Call 0871-872-9758 BOX95QU

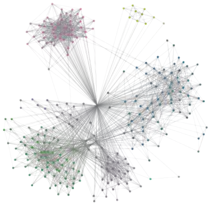
Vectorial

	name	surname	age	weight	height
0	Jorge	Perez	24	50	170
1	Pepe	Garcia	27	60	175
2	Aria	Jimenez	26	70	180
3	Maria	Ruz	25	75	181
4	Luisa	Perez	24	50	170
5	Luisa	Perez	24	50	170

Functional



Network

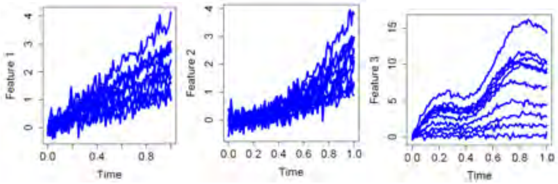


Image

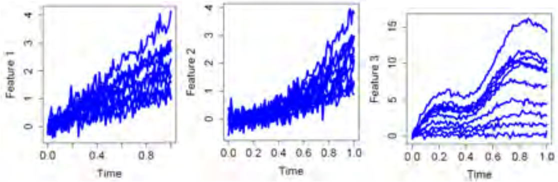
Text


ham What you doing?how are you?
ham Ok lar... Joking wif u oni...
ham dun say so early hor... U c already then say...
ham MY NO. IN LUTON 0125698789 RING ME IF UR AROUND! H*
ham Siva is in hostel aha:-.
ham Cos i was out shopping wif darren jus now n i called him 2 ask wat present
wan lor. Then he started guessing who i was wif n he finally guessed darren lor.
spam FreeMsg: Txt: CALL to No: 86888 & claim your reward of 3 hours talk time to
use from your phone now! ubscribe6GBP/ mnth inc 3hrs 16 stop?txtStop
spam Sunshine Quiz! Win a super Sony DVD recorder if you canname the capital of
Australia? Text MQUIZ to 82277. B
spam URGENT! Your Mobile No 07808726822 was awarded a L2,000 Bonus Caller Prize
02/09/03! This is our 2nd attempt to contact YOU! Call 0871-872-9758 BOX95QU

- Infinite-dimensional data.

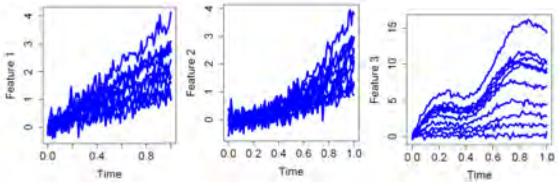


- Infinite-dimensional data.



- High interest in the literature.
  - Ferraty and Vieu [2006], Goia and Vieu [2016], Ramsay and Silverman [2002]

- Infinite-dimensional data.



- High interest in the literature.
  - Ferraty and Vieu [2006], Goia and Vieu [2016], Ramsay and Silverman [2002]
- Real world applications:

Speech recognition

Rossi and Villa [2008]

Client segmentation

Laukaitis and Rackauskas [2005]

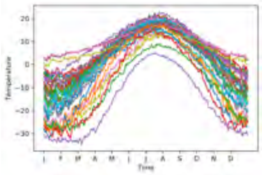
Physical processes

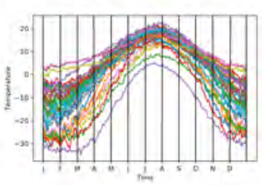
Muñoz and González [2010]

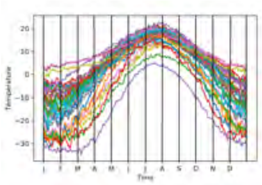
Chemical processes

Blanquero et al. [2016a,b]

Standard multivariate techniques



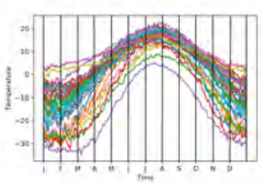

Temperature
Time
J
F
M
A
M
J
J
A
S
O
N
D



Standard multivariate techniques can be applied to *discretized* functional data

Linear Regression, SVM, Trees and RF, among others.

Standard multivariate techniques



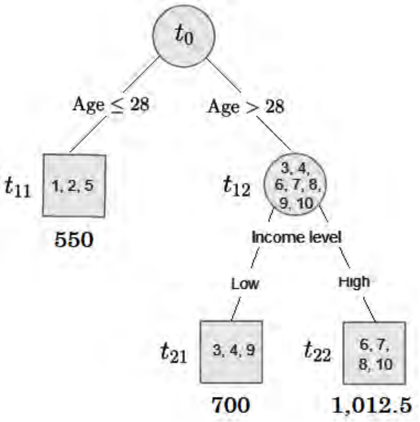
Standard multivariate techniques can be applied to *discretized* functional data

Linear Regression, SVM, Trees and RF, among others.

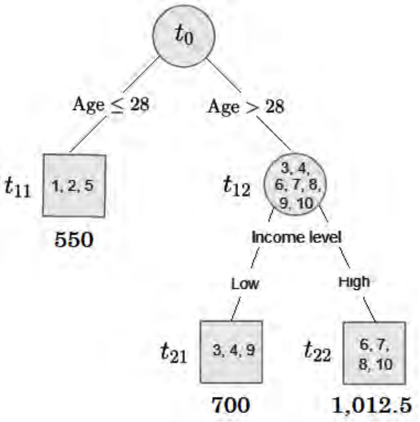
Serious drawbacks, Horváth and Kokoszka [2012]

- Curse of dimensionality, Vieu [2018].
- Bad representation of the data, Griswold et al. [2008].
- Lack of interpretation, Febrero et al. [2007]
- Intrinsic characteristics of functional data are not exploited, Borggaard and Thodberg [1992].

Individual	Age	Income level	Rent expense
1	22	Low	500
2	26	High	600
3	30	Low	700
4	32	Low	750
5	20	High	550
6	45	High	1100
7	60	High	900
8	54	High	1000
9	50	Low	650
10	48	High	1050



Individual	Age	Income level	Rent expense
1	22	Low	500
2	26	High	600
3	30	Low	700
4	32	Low	750
5	20	High	550
6	45	High	1100
7	60	High	900
8	54	High	1000
9	50	Low	650
10	48	High	1050



Popular tree methods are greedy:

CHAID, CART, C4.5 and C5.0, among others.

Literature review

Carrizosa et al. [2021], Mathematical optimization in classification and regression trees, TOP, 29(1):5-33.

Literature review

- Carrizosa et al. [2021], Mathematical optimization in classification and regression trees, TOP, 29(1):5-33.

Mixed Integer Linear Optimization (MILO)

- Aghaei et al. [2020]
- Bertsimas and Dunn [2017]
- Firat et al. [2020]
- Günlük et al. [2021]
- Hu et al. [2019], Lin et al. [2020]
- Verwer et al. [2019]
- Zantedeschi et al. [2020]
- Zhu et al. [2020]

Optimal Classification and Regression Trees

Literature review

- Carrizosa et al. [2021], Mathematical optimization in classification and regression trees, TOP, 29(1):5-33.

Mixed Integer Linear Optimization (MILO)

- Aghaei et al. [2020]
- Bertsimas and Dunn [2017]
- Firat et al. [2020]
- Günlük et al. [2021]
- Hu et al. [2019], Lin et al. [2020]
- Verwer et al. [2019]
- Zantedeschi et al. [2020]
- Zhu et al. [2020]

Constraint Programming

- Verhaeghe et al. [2019]

Literature review

- Carrizosa et al. [2021], Mathematical optimization in classification and regression trees, TOP, 29(1):5-33.

Dynamic Programming

- Demirović et al. [2020]

Literature review

- Carrizosa et al. [2021], Mathematical optimization in classification and regression trees, TOP, 29(1):5-33.

Dynamic Programming

- Demirović et al. [2020]

SAT

- Narodytska et al. [2018]
- Yu et al. [2020]

Optimal Classification and Regression Trees

Literature review

- Carrizosa et al. [2021], Mathematical optimization in classification and regression trees, TOP, 29(1):5-33.

Dynamic Programming

- Demirović et al. [2020]

SAT

- Narodytska et al. [2018]
- Yu et al. [2020]

Continuous Optimization

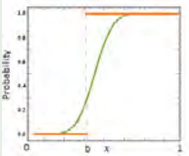
- Blanquero et al. [2021, 2020a, 2020b]

Optimal Randomized Classification and Regression Trees

Our proposal for vectorial data

Blanquero et al. [2021, 2020a, 2020b]

- We model probabilistic (as opposed to deterministic) splitting rules, defined through a CDF $F(\cdot)$:



- We seek a tradeoff between *accuracy* and *interpretability*.

Additionally:

- Local explanability.
- Counterfactual explanability.
- Cost sensitivity.
- Fairness.

Optimal Randomized Classification and Regression Trees

Aim

Adapt previous models to deal with **functional data**

Blanquero et al. [2021, 2020a, 2020b]

with two additional goals:

- Detection of *critical intervals* for prediction.
- Use of a *small number of critical intervals*, as well as a *small proportion of the curves*.

- A sample of N individuals represented by

$$\left(x_{i1}, \dots, x_{ip_1}, x_{ip_1+1}(\cdot), \dots, x_{ip_1+p_2}(\cdot), y_i\right), \quad i = 1, \dots, N,$$

where

$x_{ij} \in \mathbb{R}, \ j = 1, \dots, p_1$ *Vectorial*

$x_{ij}(\cdot) : [0,1] \to \mathbb{R}, \ j = p_1 + 1, \dots, p_1 + p_2$ *Functional*

$y_i \in \mathbb{R}$ *Response*

Jiménez-Cordero and Maldonado, 2021.

- A sample of N individuals represented by

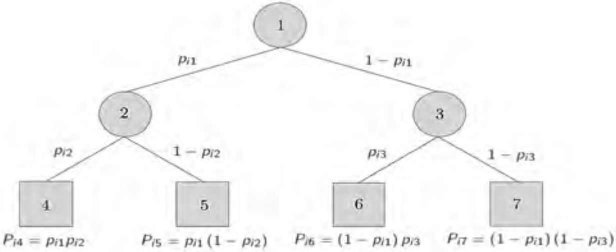
$$(x_{i1},\dots,x_{ip_1},x_{ip_1+1}(\cdot),\dots,x_{ip_1+p_2}(\cdot),y_i),\ \ i=1,\dots,N,$$

where

$x_{ij}\in\mathbb{R},\ j=1,\dots,p_1$ *Vectorial*
$x_{ij}(\cdot):[0,1]\to\mathbb{R},\ j=p_1+1,\dots,p_1+p_2$ *Functional*
$y_i\in\mathbb{R}$ *Response*

Jiménez-Cordero and Maldonado, 2021.

- A maximal binary tree of depth D, with branch τ_B and leaf τ_L nodes.



- *Oblique splits* at branch nodes $t \in \tau_B$:

- *Oblique splits* at branch nodes $t \in \tau_B$:

$$\sum_{j=1}^{p_1} a_{jt} x_{ij} \qquad + \; \mu_t, \; i = 1, \ldots, N$$

a_{jt}	coefficient of vectorial variable $j = 1, \ldots, p_1$,
μ_t	intercept,

- *Oblique splits* at branch nodes $t \in \tau_B$:

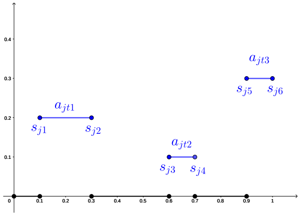
$$\sum_{j=1}^{p_1} a_{jt} x_{ij} + \sum_{j=p_1+1}^{p_1+p_2} \int_0^1 a_{jt}(s) x_{ij}(s) ds + \mu_t, \; i = 1, \ldots, N$$

a_{jt}	coefficient of vectorial variable $j = 1, \ldots, p_1$,
μ_t	intercept,
$a_{jt}(\cdot)$	coefficient of functional variable $j = p_1 + 1, \ldots, p_1 + p_2$

- *Oblique splits* at branch nodes $t \in \tau_B$:

$$\sum_{j=1}^{p_1} a_{jt} x_{ij} \; + \sum_{j=p_1+1}^{p_1+p_2} \int_0^1 a_{jt}(s) x_{ij}(s) ds \; + \; \mu_t, \; i = 1, \ldots, N$$

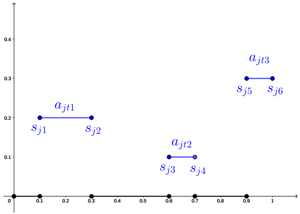
a_{jt}	coefficient of vectorial variable $j = 1, \ldots, p_1$,
μ_t	intercept,
$a_{jt}(\cdot)$	coefficient of functional variable $j = p_1 + 1, \ldots, p_1 + p_2$



- *Oblique splits* at branch nodes $t \in \tau_B$:

$$\sum_{j=1}^{p_1} a_{jt} x_{ij} \; + \sum_{j=p_1+1}^{p_1+p_2} \int_0^1 a_{jt}(s) x_{ij}(s) ds \; + \; \mu_t, \; i = 1, \ldots, N$$

a_{jt}	coefficient of vectorial variable $j = 1, \ldots, p_1$,
μ_t	intercept,
$a_{jt}(\cdot)$	coefficient of functional variable $j = p_1 + 1, \ldots, p_1 + p_2$

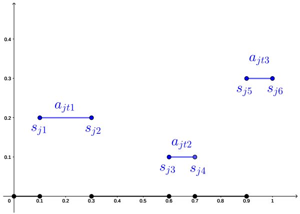


$$a_{jt}(s) = \begin{cases} 0, & \text{if } s \notin \bigcup_{k=1}^{K} [s_{j,2k-1}, s_{j,2k}] \\ \dfrac{a_{jtk}}{s_{j,2k} - s_{j,2k-1}}, & \text{if } s \in [s_{j,2k-1}, s_{j,2k}], \\ & k = 1, \ldots, K \end{cases}$$

- *Oblique splits* at branch nodes $t \in \tau_B$:

$$\sum_{j=1}^{p_1} a_{jt} x_{ij} + \sum_{j=p_1+1}^{p_1+p_2} \int_0^1 a_{jt}(s) x_{ij}(s) ds + \mu_t, \ i = 1, \dots, N$$

a_{jt}	coefficient of vectorial variable $j = 1, \dots, p_1$,
μ_t	intercept,
a_{jtk}	coefficient of functional variable $j = p_1 + 1, \dots, p_1 + p_2$ $k = 1, \dots, K$,
$[s_{j,2k-1}, s_{j,2k}]$	interval $k = 1, \dots, K, \ j = p_1 + 1, \dots, p_1 + p_2$.



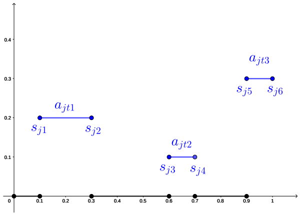
$$a_{jt}(s) = \begin{cases} 0, & \text{if } s \notin \bigcup_{k=1}^{K} [s_{j,2k-1}, s_{j,2k}] \\ \dfrac{a_{jtk}}{s_{j,2k} - s_{j,2k-1}}, & \text{if } s \in [s_{j,2k-1}, s_{j,2k}], \\ & k = 1, \dots, K \end{cases}$$

Notation

- *Oblique splits* at branch nodes $t \in \tau_B$:

$$\sum_{j=1}^{p_1} a_{jt} x_{ij} + \sum_{j=p_1+1}^{p_1+p_2} \sum_{k=1}^{K} \frac{a_{jtk}}{s_{2k} - s_{2k-1}} \int_{s_{j,2k-1}}^{s_{j,2k}} x_{ij}(s)ds + \mu_t, \ \ i = 1, \dots, N$$

a_{jt}	coefficient of vectorial variable $j = 1, \dots, p_1$,
μ_t	intercept,
a_{jtk}	coefficient of functional variable $j = p_1 + 1, \dots, p_1 + p_2$ $k = 1, \dots, K$,
$[s_{j,2k-1}, s_{j,2k}]$	interval $k = 1, \dots, K, \ j = p_1 + 1, \dots, p_1 + p_2$.



$$a_{jt}(s) = \begin{cases} 0, & \text{if } s \notin \bigcup_{k=1}^{K} [s_{j,2k-1}, s_{j,2k}] \\ \dfrac{a_{jtk}}{s_{j,2k} - s_{j,2k-1}}, & \text{if } s \in [s_{j,2k-1}, s_{j,2k}], \\ & k = 1, \dots, K \end{cases}$$

- *Probabilities* tailored to each individual $i = 1, \ldots, N$:

$$p_{it}\left(\boldsymbol{a}_t, \mu_t, \boldsymbol{s}\right) = F\left(\sum_{j=1}^{p_1} a_{jt}x_{ij} + \sum_{j=p_1+1}^{p_1+p_2}\sum_{k=1}^{K} \frac{a_{jtk}}{s_{j,2k} - s_{j,2k-1}} \int_{s_{j,2k-1}}^{s_{j,2k}} x_{ij}(s)ds + \mu_t\right), \;\; t \in \tau_B.$$

$$P_{it}\left(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{s}\right) = \prod_{t_l \in \mathcal{N}_L(t)} p_{it_l}\left(\boldsymbol{a}_{t_l}, \mu_{t_l}, \boldsymbol{s}\right) \prod_{t_r \in \mathcal{N}_R(t)} \left(1 - p_{it_r}\left(\boldsymbol{a}_{t_r}, \mu_{t_r}, \boldsymbol{s}\right)\right), \;\; t \in \tau_L.$$

- *Probabilities* tailored to each individual $i = 1, \ldots, N$:

$$p_{it}\left(\boldsymbol{a}_t, \mu_t, \boldsymbol{s}\right) = F\left(\sum_{j=1}^{p_1} a_{jt} x_{ij} + \sum_{j=p_1+1}^{p_1+p_2} \sum_{k=1}^{K} \frac{a_{jtk}}{s_{j,2k} - s_{j,2k-1}} \int_{s_{j,2k-1}}^{s_{j,2k}} x_{ij}(s) ds + \mu_t\right), \quad t \in \tau_B.$$

$$P_{it}\left(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{s}\right) = \prod_{t_l \in \mathcal{N}_L(t)} p_{it_l}\left(\boldsymbol{a}_{t_l}, \mu_{t_l}, \boldsymbol{s}\right) \prod_{t_r \in \mathcal{N}_R(t)} \left(1 - p_{it_r}\left(\boldsymbol{a}_{t_r}, \mu_{t_r}, \boldsymbol{s}\right)\right), \; t \in \tau_L.$$

- *Linear predictions* at leaf nodes $t \in \tau_L$:

$$\varphi_{it}\left(\tilde{\boldsymbol{a}}_t, \tilde{\mu}_t, \boldsymbol{s}\right) = \sum_{j=1}^{p_1} \tilde{a}_{jt} x_{ij} + \sum_{j=p_1+1}^{p_1+p_2} \sum_{k=1}^{K} \frac{\tilde{a}_{jtk}}{s_{j,2k} - s_{j,2k-1}} \int_{s_{j,2k-1}}^{s_{j,2k}} x_{ij}(s) ds + \mu_t, \; i = 1, \ldots, N$$

$\tilde{a}_{jt}$ coefficient of vectorial variable $j = 1, \ldots, p_1$,
$\tilde{a}_{jtk}$ coefficient of functional variable $j = p_1 + 1, \ldots, p_1 + p_2$, $k = 1, \ldots, K$,
$\tilde{\mu}_t$ intercept.

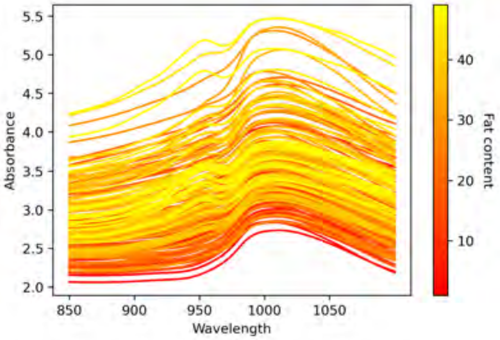
$$
\begin{aligned}
\underset{\substack{(\boldsymbol{a},\boldsymbol{\mu})\in\mathbb{R}^{(p_1+p_2K+1)|\tau_B|}\\ (\tilde{\boldsymbol{a}},\tilde{\boldsymbol{\mu}})\in\mathbb{R}^{(p_1+p_2K+1)|\tau_L|}\\ \boldsymbol{s}\in\mathbb{R}^{2p_2K}}}{\text{minimize}} \quad & \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t\in\tau_L} P_{it}\left(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{s}\right)\varphi_{it}\left(\tilde{\boldsymbol{a}}_t,\tilde{\mu}_t,\boldsymbol{s}\right)-y_i\right)^2 \text{ (MSE)}\\
& +\lambda^{\text{local}}\sum_{j=1}^{p_1+p_2}\left\|(\boldsymbol{a}_j,\tilde{\boldsymbol{a}}_j)\right\|_1 \text{ (less coefficients)}\\
& +\lambda^{\text{global}}\sum_{j=1}^{p_1+p_2}\left\|(\boldsymbol{a}_j,\tilde{\boldsymbol{a}}_j)\right\|_\infty \text{ (less variables)}\\
& +\lambda^{\text{interval}}\sum_{j=p_1+1}^{p_1+p_2}\sum_{k=1}^{K}\left\|(\boldsymbol{a}_{jk},\tilde{\boldsymbol{a}}_{jk})\right\|_\infty \text{ (less intervals)}\\
& +\lambda^{\text{length}}\sum_{j=p_1+1}^{p_1+p_2}\sum_{k=1}^{K}\left(s_{j,2k}-s_{j,2k-1}\right) \text{ (less proportion of curves)}\\
\text{s.t.} \quad & s_{j,k}\leq s_{j,k+1},\ j=p_1+1,\ldots,p_1+p_2,\ k=1,\ldots,K-1
\end{aligned}
$$

The model

$$
\begin{aligned}
\underset{\substack{(\boldsymbol{a},\boldsymbol{\mu})\in\mathbb{R}^{(p_1+p_2K+1)|\tau_B|}\\ (\tilde{\boldsymbol{a}},\tilde{\boldsymbol{\mu}})\in\mathbb{R}^{(p_1+p_2K+1)|\tau_L|}\\ \boldsymbol{s}\in\mathbb{R}^{2p_2K}}}{\text{minimize}} \quad & \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t\in\tau_L} P_{it}\left(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{s}\right)\varphi_{it}\left(\tilde{\boldsymbol{a}}_t,\tilde{\mu}_t,\boldsymbol{s}\right) - y_i\right)^2 \text{ (MSE)}\\
& + \lambda^{\text{local}}\sum_{j=1}^{p_1+p_2}\left\|\left(\boldsymbol{a}_j,\tilde{\boldsymbol{a}}_j\right)\right\|_1 \text{ (less coefficients)}\\
& + \lambda^{\text{global}}\sum_{j=1}^{p_1+p_2}\left\|\left(\boldsymbol{a}_j,\tilde{\boldsymbol{a}}_j\right)\right\|_\infty \text{ (less variables)}\\
& + \lambda^{\text{interval}}\sum_{j=p_1+1}^{p_1+p_2}\sum_{k=1}^{K}\left\|\left(\boldsymbol{a}_{jk},\tilde{\boldsymbol{a}}_{jk}\right)\right\|_\infty \text{ (less intervals)}\\
& + \lambda^{\text{length}}\sum_{j=p_1+1}^{p_1+p_2}\sum_{k=1}^{K}\left(s_{j,2k}-s_{j,2k-1}\right) \text{ (less proportion of curves)}\\
\text{s.t.} \quad & s_{j,k}\le s_{j,k+1},\ j=p_1+1,\dots,p_1+p_2,\ k=1,\dots,K-1
\end{aligned}
$$

- There exists an equivalent nonlinear smooth formulation with linear constraints.

- Tecator data set: the near-infrared absorbance spectra of 215 samples of finely chopped pork. The response variable is the fat content.



Implementation details

- Logistic CDF.
- SLSQP solver in the `scipy.optimize` package in Python 3.7.
- Intel® Core™ i7-8550U CPU 1.80GHz with 8 GB RAM.

Implementation details

- Logistic CDF.
- SLSQP solver in the `scipy.optimize` package in Python 3.7.
- Intel® Core™ i7-8550U CPU 1.80GHz with 8 GB RAM.

Comparison using 10-fold cross-validation

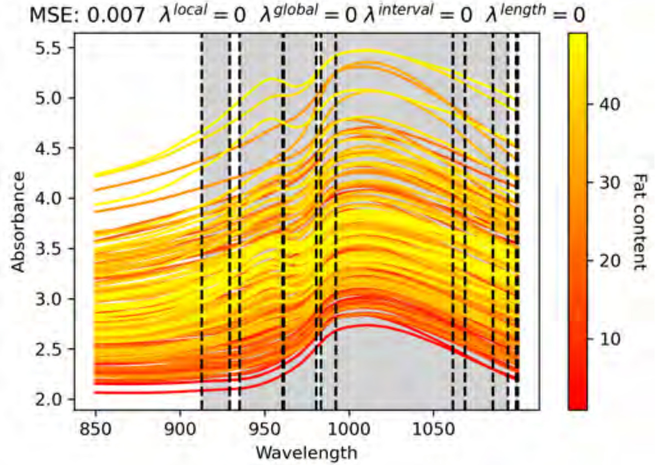
Optimal Randomized Regression Trees with $D = 1$ and $\lambda^{\text{local}} = \lambda^{\text{global}} = \lambda^{\text{interval}} = \lambda^{\text{length}} = 0$, with benchmarks:

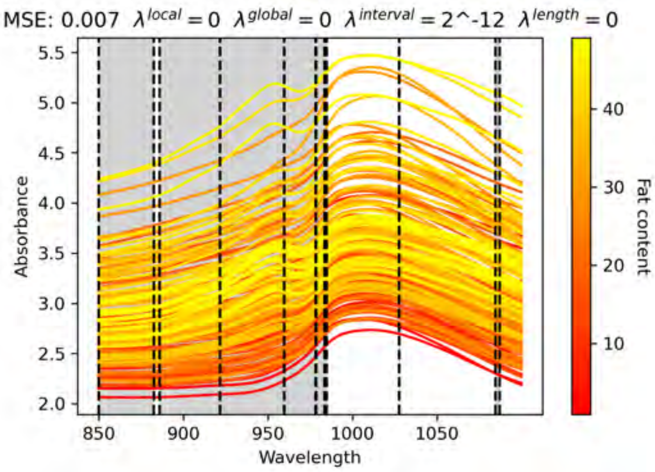
- Random Forest in Breiman [2001]
- SVR-FD in Blanquero et al. [2020]

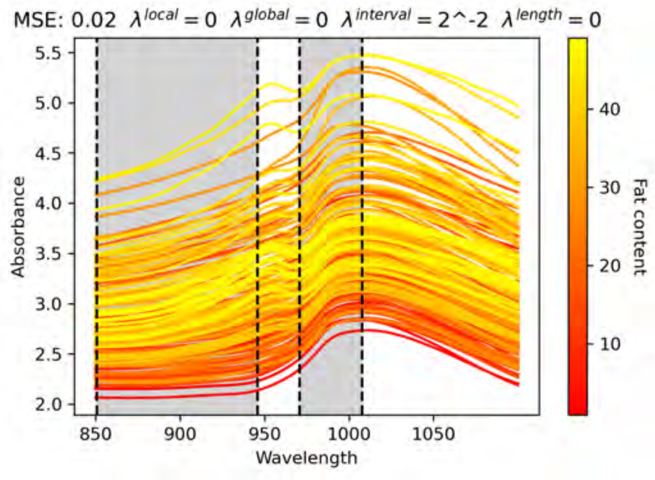
Table: Results for ORRT with $D = 1$ and $\lambda^{\text{local}} = \lambda^{\text{global}} = \lambda^{\text{interval}} = \lambda^{\text{length}} = 0$.

K	Out-of-sample average SSR		
	RF	ORRT $D = 1$	SVR-FD
	2.13		
1		4.74	0.25
2		0.54	0.23
3		0.22	0.24
4		0.18	0.29
5		0.19	0.41
6		0.17	0.40
7		0.13	0.45
8		0.18	0.45
9		0.19	0.49
10		0.19	0.50

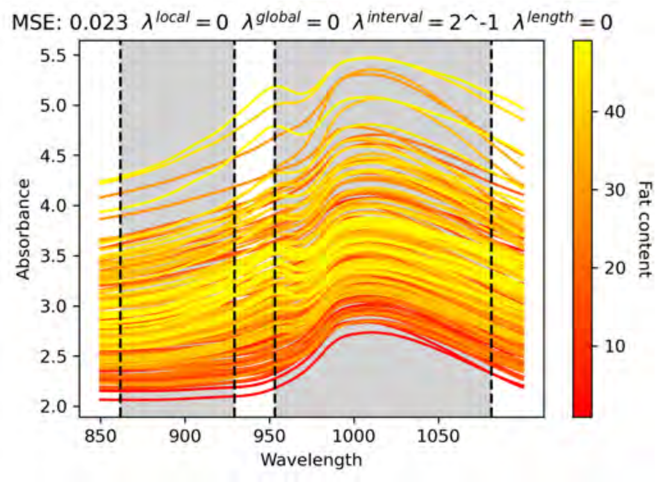
Controlling the number of critical intervals


MSE: 0.007 $\lambda^{local} = 0$ $\lambda^{global} = 0$ $\lambda^{interval} = 0$ $\lambda^{length} = 0$
Absorbance
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
850
900
950
1000
1050
Wavelength
Fat content
40
30
20
10

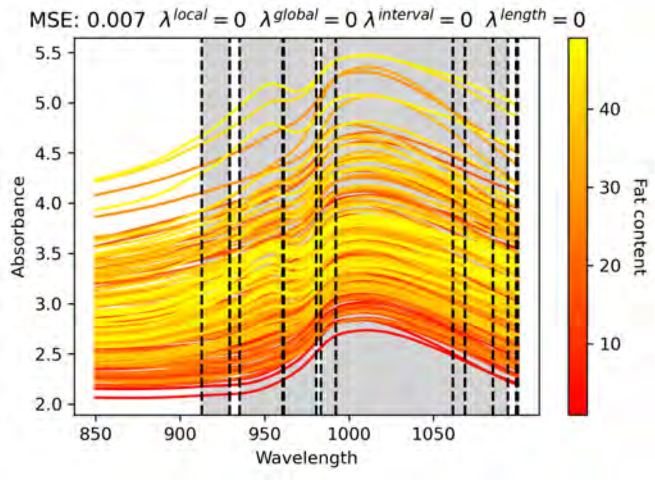

MSE: 0.007 λ^local = 0 λ^global = 0 λ^interval = 2^-12 λ^length = 0
Absorbance
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
850
900
950
1000
1050
Wavelength
Fat content
40
30
20
10

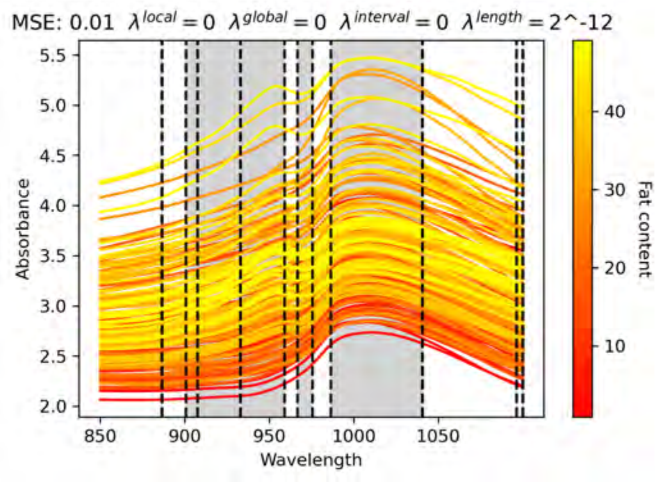


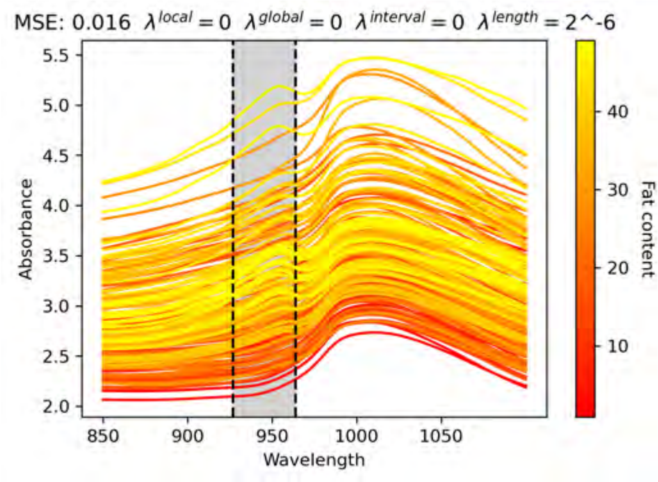
MSE: 0.02 $\lambda^{local} = 0$ $\lambda^{global} = 0$ $\lambda^{interval} = 2$^-2 $\lambda^{length} = 0$
Absorbance
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
850
900
950
1000
1050
Wavelength
40
30
20
10
Fat content


MSE: 0.023 λ^local = 0 λ^global = 0 λ^interval = 2^-1 λ^length = 0
Absorbance
Wavelength
Fat content
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
850
900
950
1000
1050
40
30
20
10

Controlling the proportion of curve


MSE: 0.007 $\lambda^{local} = 0$ $\lambda^{global} = 0$ $\lambda^{interval} = 0$ $\lambda^{length} = 0$
Absorbance
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
850
900
950
1000
1050
Wavelength
Fat content
40
30
20
10


MSE: 0.01 λ^local = 0 λ^global = 0 λ^interval = 0 λ^length = 2^-12
Absorbance
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
850
900
950
1000
1050
Wavelength
40
30
20
10
Fat content


MSE: 0.016 λ^local = 0 λ^global = 0 λ^interval = 0 λ^length = 2^-6
Absorbance
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
850
900
950
1000
1050
Wavelength
40
30
20
10
Fat content

Conclusions

- Optimal Randomized Classification and Regression Trees, a *continuous optimization* approach to build trees with *probabilistic* cuts, has been adapted to handle *functional data*.
- *Critical intervals* for prediction are detected simultaneously.
- The *proportion of the curves* to be used, as well as the *number of critical intervals* can be controlled.
- Preliminary results indicate that our approach is *competitive to benchmark methods, including RF.*

Conclusions

- Optimal Randomized Classification and Regression Trees, a *continuous optimization* approach to build trees with *probabilistic* cuts, has been adapted to handle *functional data*.
- *Critical intervals* for prediction are detected simultaneously.
- The *proportion of the curves* to be used, as well as the *number of critical intervals* can be controlled.
- Preliminary results indicate that our approach is *competitive to benchmark methods, including RF.*

Future research

- Adapt our approach to deal with other complex data, for instance, *text data* or *network data*.
- Embed our approach in *bagging* (as RF) or *boosting ensembles*.

Bibliography I

S. Aghaei, M.J. Azizi, and P. Vayanos.
Learning optimal and fair decision trees for non-discriminative decision-making.
In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1418–1426, 2019.

S. Aghaei, A. Gomez, and P. Vayanos.
Learning optimal classification trees: Strong max-flow formulations.
arXiv preprint arXiv:2002.09142, 2020.

D. Bertsimas and J. Dunn.
Optimal classification trees.
Machine Learning, 106(7):1039–1082, 2017.

R. Blanquero, E. Carrizosa, O. Chis, N. Esteban, A. Jiménez-Cordero, J. F. Rodríguez, and M. R. Sillero-Denamiel.
On extreme concentrations in chemical reaction networks with incomplete measurements.
Industrial & Engineering Chemistry Research, 55(44):11417–11430, 2016a.

R. Blanquero, E. Carrizosa, A. Jiménez-Cordero, and J. F. Rodríguez.
A global optimization method for model selection in chemical reactions networks.
Computers & Chemical Engineering, 93:52–62, 2016b.

R. Blanquero, E. Carrizosa, A. Jiménez-Cordero, and B. Martín-Barragán.
Selection of time instants and intervals with support vector regression for multivariate functional data.
Computers & Operations Research, 123:105050, 2020.

R. Blanquero, E. Carrizosa, C. Molero-Río, and D. Romero Morales.
Optimal randomized classification trees.
Computers & Operations Research, 132:105281, 2021.

R. Blanquero, E. Carrizosa, C. Molero-Río, and D. Romero Morales.
Sparsity in optimal randomized classification trees.
European Journal of Operational Research, 284(1):255 – 272, 2020.

R. Blanquero, E. Carrizosa, C. Molero-Río, and D. Romero Morales.
On sparse optimal regression trees.
Technical report, IMUS, Sevilla, Spain,
https://www.researchgate.net/publication/341099512_On_Sparse_Optimal_Regression_Trees,
2020.

C. Borggaard and H. H. Thodberg.
Optimal minimal neural interpretation of spectra.
Analytical chemistry, 64(5):545–551, 1992.

L. Breiman, J.H. Friedmann, R.A. Olshen, and C.J. Stone.
Classification and Regression Trees.
Wadsworth, Belmont, CA, 1984.

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

E. Carrizosa, C. Molero-Río, and D. Romero Morales.
Mathematical optimization in classification and regression trees.
TOP, 29(1):5–33, 2021.

E. Demirović, A. Lukina, E. Hebrard, J. Chan, J. Bailey, C. Leckie, K. Ramamohanarao, and P.J. Stuckey.
MurTree: Optimal Classification Trees via Dynamic Programming and Search.
arXiv preprint arXiv:2007.12652, 2020.

J. Dunn.
Optimal trees for prediction and prescription.
PhD thesis, Massachusetts Institute of Technology, 2018.

M. Febrero, P. Galeano, and W. González-Manteiga.
A functional analysis of nox levels: location and scale estimation and outlier detection.
Computational Statistics, 22(3):411–427, 2007.

F. Ferraty and P. Vieu.
Nonparametric functional data analysis: theory and practice.
Springer Science & Business Media, 2006.

M. Firat, G. Crognier, A.F. Gabor, C.A.J. Hurkens, and Y. Zhang.
Column generation based heuristic for learning classification trees.
Computers & Operations Research, 116:104866, 2020.

C. Gambella, B. Ghaddar, and J. Naoum-Sawaya.
Optimization models for machine learning: A survey.
European Journal of Operational Research, 290(3):807–828, 2021.

A. Goia and P. Vieu.
An introduction to recent advances in high/infinite dimensional statistics, 2016.

C. K. Griswold, R. Gomulkiewicz, and N. Heckman.
Hypothesis testing in comparative and experimental studies of function-valued traits.
Evolution, 62(5):1229–1242, 2008.

O. Günlük, J. Kalagnanam, M. Menickelly, and K. Scheinberg.
Optimal decision trees for categorical data via integer programming. *Forthcoming in* Journal of Global Optimization.
https://doi.org/10.1007/s10898-021-01009-y, 2021.

L. Horváth and P. Kokoszka.
Inference for functional data with applications, volume 200.
Springer Science & Business Media, 2012.

X. Hu, C. Rudin, and M. Seltzer.
Optimal sparse decision trees.
In *Advances in Neural Information Processing Systems 32*, pages 7265–7273. 2019.

M. A. Jiménez Cordero.
Classification and regression with functional data: a mathematical optimization approach.
PhD thesis, University of Seville, 2019.

A. Jiménez-Cordero and S. Maldonado.
Automatic feature scaling and selection for support vector machine classification with functional data.
Applied Intelligence, 51(1):161–184, 2021.

A. Laukaitis and A. Račkauskas.
Functional data analysis for clients segmentation tasks.
European journal of operational research, 163(1):210–216, 2005.

J. Lin, C. Zhong, D. Hu, C. Rudin, and M. Seltzer.
Generalized and scalable optimal sparse decision trees.
arXiv preprint arXiv:2006.08690, 2020.

A. Muñoz and J. González.
Representing functional data using support vector machines.
Pattern Recognition Letters, 31(6):511–516, 2010.

N. Narodytska, A. Ignatiev, F. Pereira, and J. Marques-Silva.
Learning Optimal Decision Trees with SAT.
In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, pages 1362–1368, 2018.

J. O. Ramsay and B. W. Silverman.
Applied functional data analysis: methods and case studies, volume 77.
Springer, 2002.

F. Rossi and N. Villa.
Recent advances in the use of svm for functional data classification.
Functional and operatorial statistics, pages 273–280, 2008.

H. Verhaeghe, S. Nijssen, G. Pesant, C.-G. Quimper, and P. Schaus.
Learning optimal decision trees using constraint programming.
In *The 25th International Conference on Principles and Practice of Constraint Programming (CP2019)*, 2019.

S. Verwer, Y. Zhang, and Q.C. Ye.
Learning optimal classification trees using a binary linear program formulation.
In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1625–1632, 2019.

P. Vieu.
On dimension reduction models for functional data.
Statistics & Probability Letters, 136:134–138, 2018.

J. Yu, A. Ignatiev, P.J. Stuckey, and P. Le Bodic.
Computing Optimal Decision Sets with SAT.
arXiv preprint arXiv:2007.15140, 2020.

V. Zantedeschi, M.J. Kusner, and V. Niculae.
Learning binary trees via sparse relaxation.
arXiv preprint arXiv:2010.04627, 2020.

H. Zhu, P. Murali, D.T. Phan, L.M. Nguyen, and J.R. Kalagnanam.
A scalable MIP-based method for learning optimal multivariate decision trees.
Advances in Neural Information Processing Systems, 33, 2020.

Thank you for your attention!

mmolero@us.es

https://www.researchgate.net/profile/Cristina_Molero-Rio