Mathematical Optimization in Data Science

Emilio Carrizosa
Instituto de Matemáticas de la Universidad de Sevilla
Lleida, 5th July 2019

www.grupo.us.es/gioptim

instituto de matemáticas universidad de sevilla

www.riseneeds.eu

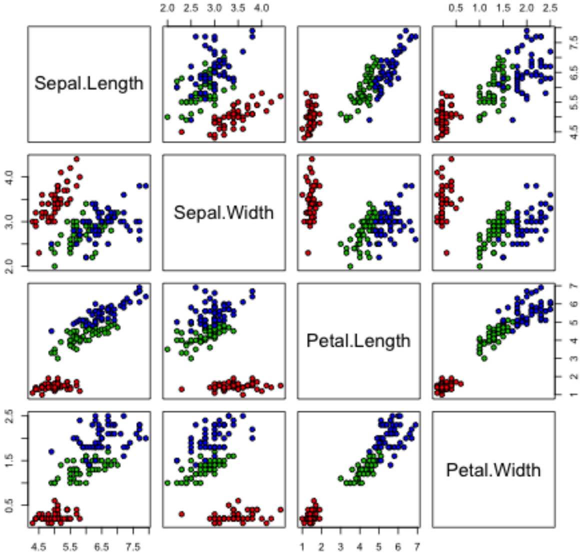
@ needs_project

- Sandra Benítez-Peña
- Rafael Blanquero
- Vanesa Guerrero
- M. Asunción Jiménez-Cordero
- Belén Martín-Barragán
- Cristina Molero-Río
- Alba V Olivares-Nadal
- Pepa Ramírez-Cobo
- Dolores Romero Morales
- Remedios Sillero-Denamiel

NeEDS Network of European Data Scientists www.imus.us.es/eps-data-optimization

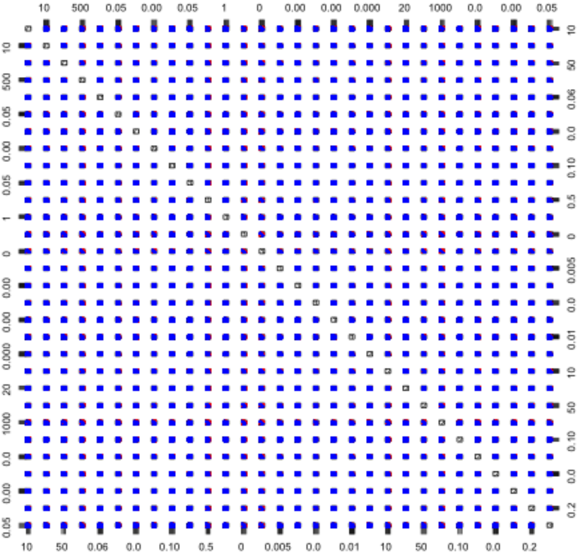
Visualization

Iris Data



Sepal.Length
Sepal.Width
Petal.Length
Petal.Width
2.0 2.5 3.0 3.5 4.0
0.5 1.0 1.5 2.0 2.5
4.5 5.5 6.5 7.5
1 2 3 4 5 6 7

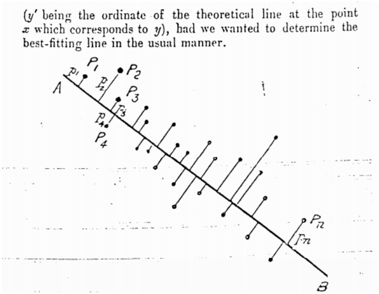
Winsconsin Breast Cancer Data



PCA

- Principal Component Analysis (PCA): way of projecting properly a data set $\subset \mathbb{R}^d$ into an affine space of smaller dimension

Pearson. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, 1901


(y' being the ordinate of the theoretical line at the point x which corresponds to y), had we wanted to determine the best-fitting line in the usual manner.

PCA

- We're given $\{u_1, \ldots, u_N\} \subset \mathbb{R}^d$, wlog, $\frac{1}{N}\sum_{i=1}^{N} u_i = 0_d$

PCA

- We're given $\{u_1, \ldots, u_N\} \subset \mathbb{R}^d$, wlog, $\frac{1}{N}\sum_{i=1}^{N} u_i = 0_d$
- Seeking orthonormal $c_1, \ldots, c_k$ s.t.
  $u_i \approx \pi_{\{c_1,\ldots,c_k\}}(u_i) \qquad \forall i = 1, 2, \ldots, N :$

$$\min_{c_1,\ldots,c_k:\ \text{orthonormal}} \frac{1}{N}\sum_{i=1}^{N} \left\| u_i - \pi_{\{c_1,\ldots,c_k\}}(u_i) \right\|^2$$

PCA

- We're given $\{u_1, \dots, u_N\} \subset \mathbb{R}^d$, wlog, $\frac{1}{N}\sum_{i=1}^{N} u_i = 0_d$
- Seeking orthonormal $c_1, \dots, c_k$ s.t.
  $u_i \approx \pi_{\{c_1,\dots,c_k\}}(u_i) \qquad \forall i = 1, 2, \dots, N :$

$$\min_{c_1,\dots,c_k:\text{ orthonormal}} \frac{1}{N}\sum_{i=1}^{N} \left\| u_i - \pi_{\{c_1,\dots,c_k\}}(u_i) \right\|^2$$

- $V := \frac{1}{N}(u_1|u_2|\dots|u_N)\cdot(u_1|u_2|\dots|u_N)^\top$ (covariance matrix), an sdp matrix
- Problem equivalent to

$$\begin{array}{ll} \min & \frac{1}{N}\sum_{i=1}^{N}\|u_i\|^2 - \frac{1}{N}\sum_{j=1}^{k} c_j^\top \cdot V \cdot c_j \\ & c_i^\top c_j = \delta_{ij} \qquad \forall i, j = 1 \dots k \end{array}$$

PCA

- We're given $\{u_1, \ldots, u_N\} \subset \mathbb{R}^d$, wlog, $\frac{1}{N}\sum_{i=1}^{N} u_i = 0_d$
- Seeking orthonormal $c_1, \ldots, c_k$ s.t.
  $u_i \approx \pi_{\{c_1,\ldots,c_k\}}(u_i) \qquad \forall i = 1, 2, \ldots, N:$

$$\min_{c_1,\ldots,c_k:\ \text{orthonormal}} \frac{1}{N}\sum_{i=1}^{N} \left\|u_i - \pi_{\{c_1,\ldots,c_k\}}(u_i)\right\|^2$$

- $V := \frac{1}{N}(u_1|u_2|\ldots|u_N) \cdot (u_1|u_2|\ldots|u_N)^\top$ (covariance matrix), an sdp matrix
- Problem equivalent to

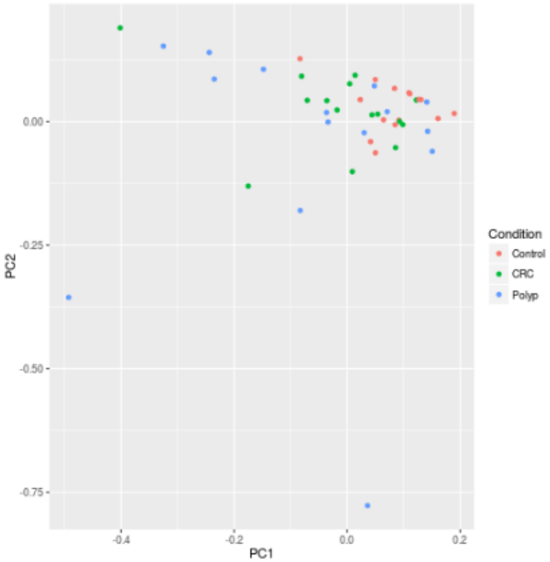
$$\frac{1}{N}\sum_{i=1}^{N}\|u_i\|^2 \quad - \quad \begin{array}{l} \max \frac{1}{N}\sum_{j=1}^{k} c_j^\top \cdot V \cdot c_j \\ c_i^\top c_j = \delta_{ij} \qquad \forall i,j = 1\ldots k \end{array}$$

PCA

$$\begin{array}{ll} \min & \frac{1}{N}\sum_{i=1}^{N}\|u_i\|^2 - \frac{1}{N}\sum_{j=1}^{k} c_j^\top \cdot V \cdot c_j \\ & c_i^\top c_j = \delta_{ij} \qquad \forall i,j = 1\ldots k \end{array}$$

Calculating principal components

- Optimal $c_1, c_2, \ldots, c_k$: unit eigenvectors associated with the k largest eigenvalues of the sdp matrix V



0.00
-0.25
-0.50
-0.75
PC2
-0.4
-0.2
0.0
0.2
PC1
Condition
Control
CRC
Polyp

Sparse PCA. A few references

- d'Aspremont, A., El Ghaoui, L., Jordan, M., and Lanckriet, G. "A Direct Formulation for Sparse PCA Using Semidefinite Programming", *SIAM Review*, 2007.
- Carrizosa, E., and Guerrero, V. "Biobjective Sparse Principal Component Analysis". *Journal of Multivariate Analysis*, 2014.
- Carrizosa, E., and Guerrero, V. "rs-Sparse principal component analysis: A mixed integer nonlinear programming approach with VNS". *Computers & Operations Research*, 2014.
- Jolliffe, I.T., N. T. Trendafilov and M. Uddin. "A Modified Principal Component Technique Based on the LASSO", *J. of Computational and Graphical Statistics*, 2003.
- Krauthgamer, R., Nadler, .B., and Vilenchik, D. "Do semidefinite relaxations solve sparse PCA up to the information limit?", *The Annals of Statistics*, 2015
- Ma, Z. "Sparse principal component analysis and iterative thresholding". *Annals of Statistics*, 2013.
- McCabe, G. P. "Principal Variables", *Technometrics*, 26, 1984.
- Vines, S. K. "Simple Principal Components", *Applied Statistics*, 2000.
- Zou, H., T. Hastie and R. Tibshirani. "Sparse Principal Component Analysis", *J. of Computational and Graphical Statistics*, 2006.

rs-Sparse PCA. Carrizosa and Guerrero-Lozano, 2014

PCA

$$\begin{array}{ll} \min & \frac{1}{N}\sum_{i=1}^{N} \left\| u_i - \pi_{\{c_1,\ldots,c_k\}}(u_i) \right\|^2 \\ & c_1,\ldots,c_k : \text{ orthonormal} \end{array}$$

Sparse PCA

$$
\begin{array}{ll}
\min & \frac{1}{N}\sum_{i=1}^{N} \left\| u_i - \pi_{\{c_1,\ldots,c_k\}}(u_i) \right\|^2 \\
 & c_1,\ldots,c_k : \text{ orthonormal} \\
 & \quad + \text{ global sparsity constraints:}
\end{array}
$$

Sparse PCA

$$\begin{array}{ll} \min & \frac{1}{N}\sum_{i=1}^{N} \left\| u_i - \pi_{\{c_1,\ldots,c_k\}}(u_i) \right\|^2 \\ & c_1,\ldots,c_k : \text{ orthonormal} \\ & \;+\text{ global sparsity constraints:} \end{array}$$

Global sparsity constraints

1. Each variable is nonzero in at most r components c_j

Hard constraints

rs-Sparse PCA. Carrizosa and Guerrero-Lozano, 2014

Sparse PCA

$$\begin{array}{ll} \min & \frac{1}{N}\sum_{i=1}^{N} \left\| u_i - \pi_{\{c_1,\ldots,c_k\}}(u_i) \right\|^2 \\ & c_1, \ldots, c_k : \text{ orthonormal} \\ & + \text{ global sparsity constraints:} \end{array}$$

Global sparsity constraints

1. Each variable is nonzero in at most r components c_j
2. Each c_j has at most s nonzero elements

Hard constraints

Define: $z_{il} = \begin{cases} 1 & \text{if} \quad c_{il} \neq 0 \\ 0 & \text{else} \end{cases} \quad i = 1 \ldots k, l = 1 \ldots d$

rs-Sparse PCA. MINLP formulation

Define: $z_{il} = \begin{cases} 1 & \text{if} \quad c_{il} \neq 0 \\ 0 & \text{else} \end{cases} \quad i = 1 \ldots k, l = 1 \ldots d$

$$|c_{il}| \leq z_{il} \qquad \forall i, l$$

rs-Sparse PCA. MINLP formulation

Define: $z_{il} = \begin{cases} 1 & \text{if} \quad c_{il} \neq 0 \\ 0 & \text{else} \end{cases} \quad i = 1 \ldots k, l = 1 \ldots d$

$$|c_{il}| \leq z_{il} \qquad \forall i, l$$

$$\sum_{i=1}^{k} z_{il} \leq r \quad \forall l = 1 \ldots d$$

rs-Sparse PCA. MINLP formulation

Define: $z_{il} = \begin{cases} 1 & \text{if} \quad c_{il} \neq 0 \\ 0 & \text{else} \end{cases} \quad i = 1 \ldots k, l = 1 \ldots d$

$$|c_{il}| \leq z_{il} \qquad \forall i, l$$

$$\sum_{i=1}^{k} z_{il} \leq r \quad \forall l = 1 \ldots d$$

$$\sum_{l=1}^{d} z_{il} \leq s \quad \forall i = 1 \ldots k$$

rs-Sparse PCA. MINLP formulation

$$
\begin{array}{lll}
\min & \frac{1}{N}\sum_{i=1}^{N} \left\| u_i - \pi_{\{c_1,\ldots,c_k\}}(u_i) \right\|^2 & \\
& c_i^\top c_j = \delta_{ij} & \forall i, j \\
& |c_{il}| \le z_{il} & \forall i, l \\
& \sum_{i=1}^{k} z_{il} \le r & \forall l = 1 \ldots d \\
& \sum_{l=1}^{n} z_{il} \le s & \forall i = 1 \ldots k \\
& z_{il} \in \{0, 1\} & \forall i, l
\end{array}
$$

rs-Sparse PCA. MINLP formulation

$$
\begin{array}{lll}
\max & \sum_{j=1}^{k} c_j^\top \cdot V \cdot c_j & \\
& c_i^\top c_j = \delta_{ij} & \forall i, j \\
& |c_{il}| \leq z_{il} & \forall i, l \\
& \sum_{i=1}^{k} z_{il} \leq r & \forall l = 1 \dots d \\
& \sum_{l=1}^{n} z_{il} \leq s & \forall i = 1 \dots k \\
& z_{il} \in \{0, 1\} & \forall i, l
\end{array}
$$

1s-Sparse PCA

$$
\begin{array}{lll}
\max & \sum_{j=1}^{k} c_j^\top \cdot V \cdot c_j & \\
 & {\color{red} c_i^\top c_j = \delta_{ij}} & \forall i, j \\
 & |c_{il}| \leq z_{il} & \forall i, l \\
 & \sum_{i=1}^{k} z_{il} = 1 & \forall l = 1 \dots d \\
 & \sum_{l=1}^{n} z_{il} = s & \forall i = 1 \dots k \\
 & z_{il} \in \{0, 1\} & s \forall i, l
\end{array}
$$

1s-Sparse PCA

$$
\begin{array}{lll}
\max & \sum_{j=1}^{k} c_j^\top \cdot V \cdot c_j & \\
 & c_i^\top c_i = 1 & \forall i \\
 & |c_{il}| \leq z_{il} & \forall i, l \\
 & \sum_{i=1}^{k} z_{il} = 1 & \forall l = 1 \ldots d \\
 & \sum_{l=1}^{n} z_{il} = s & \forall i = 1 \ldots k \\
 & z_{il} \in \{0, 1\} & s\forall i, l
\end{array}
$$

Fixing z ...

$$
\begin{array}{lll}
\max & \sum_{j=1}^{k} c_j^\top \cdot V \cdot c_j & \\
& c_i^\top c_i = 1 & \forall i \\
& |c_{il}| \leq z_{il} & \forall i, l \\
& \sum_{i=1}^{k} z_{il} = 1 & \forall l = 1 \dots d \\
& \sum_{l=1}^{n} z_{il} = s & \forall i = 1 \dots k \\
& z_{il} \in \{0, 1\} & \forall i, l
\end{array}
$$

Fixing z ...

$$
\begin{array}{lll}
\max & \sum_{j=1}^{k} c_j^\top \cdot V \cdot c_j & \\
 & c_i^\top c_i = 1 & \forall i \\
 & c_{il} = 0 & \forall i, l : z_{il} = 0
\end{array}
$$

Fixing z ...

$$
\begin{array}{lll}
\max & \sum_{j=1}^{k} c_j^\top \cdot V \cdot c_j & \\
& c_i^\top c_i = 1 & \forall i \\
& c_{il} = 0 & \forall i, l : z_{il} = 0
\end{array}
$$

Resulting problem ...

- Separable in k problems (of classical PCA-type)

Fixing z ...

$$\begin{array}{lll} \max & \sum_{j=1}^{k} c_j^\top \cdot V \cdot c_j & \\ & c_i^\top c_i = 1 & \forall i \\ & c_{il} = 0 & \forall i, l : z_{il} = 0 \end{array}$$

Resulting problem ...

- Separable in k problems (of classical PCA-type)
- Amounts to solving largest eigenvalue and associated eigenvector of k submatrices of V

A heuristic

1. "Judiciously" choose z
2. Find the optimal c of z fixed (by calculating k eigenvalues and eigenvectors)

A heuristic

1. "Judiciously" choose z
2. Find the optimal c of z fixed (by calculating k eigenvalues and eigenvectors)

Choosing z

- Easily available: $c_1^*, \ldots, c_k^*$, principal components
- Controlled rounding of $c_1^*, \ldots, c_k^*$:

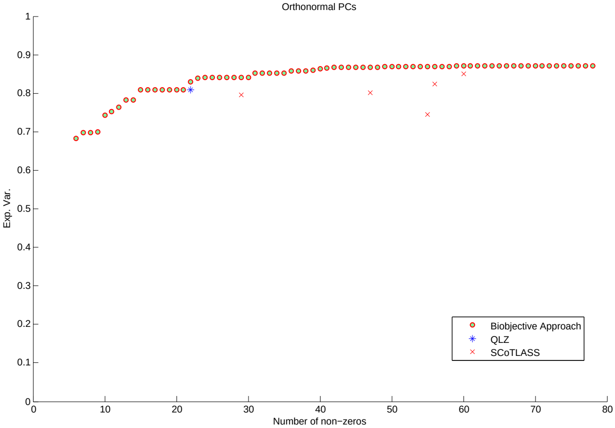
A heuristic

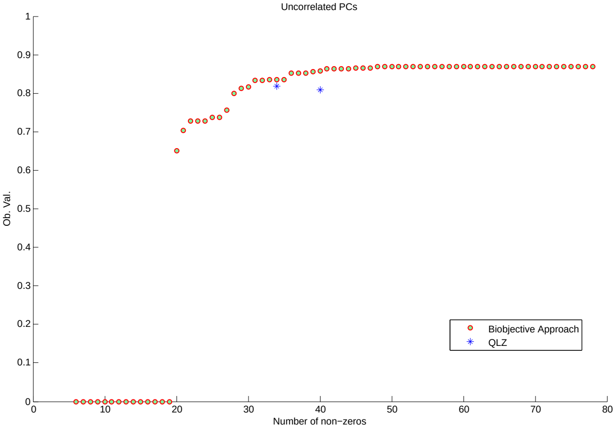
1. "Judiciously" choose z
2. Find the optimal c of z fixed (by calculating k eigenvalues and eigenvectors)

Choosing z

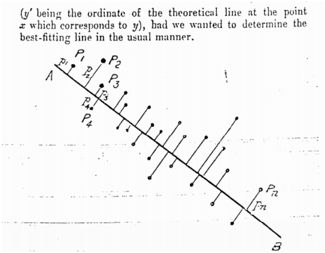
- Easily available: $c_1^*, \dots, c_k^*$, principal components
- Controlled rounding of $c_1^*, \dots, c_k^*$:

$$
\begin{array}{lll}
\max & \sum_{i=1}^{d} \sum_{l=1}^{k} |c_{il}^*| z_{il} & \\
 & \sum_{l=1}^{k} z_{il} = 1 & \forall i = 1, \dots, d \\
 & \sum_{i=1}^{n} z_{il} \leq s & \forall l = 1, \dots, k \\
 & \sum_{i=1}^{n} z_{il} \geq 1 & \forall l = 1, \dots, k \\
 & z_{il} \geq 0 & \forall i, l
\end{array}
$$


Orthonormal PCs
Exp. Var.
Number of non−zeros
Biobjective Approach
QLZ
SCoTLASS
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80

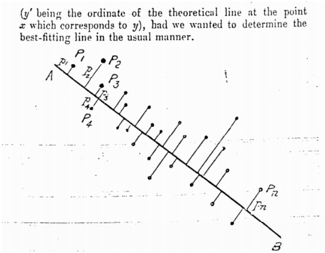

Uncorrelated PCs
Ob. Val.
Number of non−zeros
Biobjective Approach
QLZ
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80

The limits of PCA

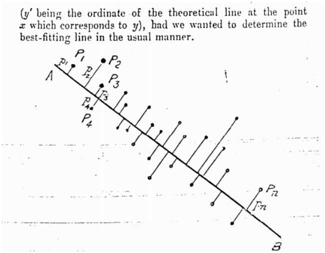

(y' being the ordinate of the theoretical line at the point x which corresponds to y), had we wanted to determine the best-fitting line in the usual manner.

- Input: data coordinates in $\mathbb{R}^d$

The limits of PCA

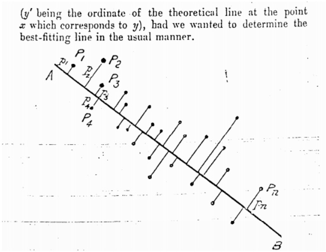


- Input: data coordinates in $\mathbb{R}^d$
- PCA output may not properly reflect proximities

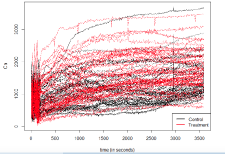


- Input: data coordinates in $\mathbb{R}^d$
- PCA output may not properly reflect proximities
- Not (directly applicable) when e.g.
  - Coordinates missing
  - Input is not Euclidean, and a disimilarity is given instead

The limits of PCA



- Input: data coordinates in $\mathbb{R}^d$
- PCA output may not properly reflect proximities
- Not (directly applicable) when e.g.
  - Coordinates missing
  - Input is not Euclidean, and a disimilarity is given instead



Kruskal. Psychometrika, 1964

V :	$v_1, v_2, \ldots, v_N$
$\boldsymbol{\delta}$:	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$

Kruskal. Psychometrika, 1964

set of N individuals

V :	$v_1, v_2, \ldots, v_N$
$\boldsymbol{\delta}$:	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$

MultiDimensional Scaling

Kruskal. Psychometrika, 1964

set of N individuals

dissimilarities

$V:$	$v_1, v_2, \ldots, v_N$
$\boldsymbol{\delta}:$	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$

Kruskal. Psychometrika, 1964

set of N individuals

dissimilarities

V :	$v_1, v_2, \ldots, v_N$
$\boldsymbol{\delta}$:	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$

- $v_i \longmapsto \boldsymbol{c}_i \in \mathbb{R}^n$

Kruskal. Psychometrika, 1964

set of N individuals	$V:$	$v_1, v_2, \ldots, v_N$
dissimilarities	$\boldsymbol{\delta}:$	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$

- $v_i \longmapsto \boldsymbol{c}_i \in \mathbb{R}^n$
- $\|\boldsymbol{c}_i - \boldsymbol{c}_j\| \delta_{ij} \; \forall i, j$

MultiDimensional Scaling

Kruskal. Psychometrika, 1964

set of N individuals	V :	$v_1, v_2, \ldots, v_N$
dissimilarities	$\boldsymbol{\delta}$:	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$

- $v_i \longmapsto \boldsymbol{c}_i \in \mathbb{R}^n$
- $\|\boldsymbol{c}_i - \boldsymbol{c}_j\| \delta_{ij} \; \forall i, j$
- $\min_{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_N} \sum_{i,j} \left(\|\boldsymbol{c}_i - \boldsymbol{c}_j\|^2 \delta_{ij}^2 \right)^2$
  - Unconstrained optimization with smooth function
  - Highly multimodal when δ stongly violates triangle inequality

MDS with objects

$V:$	$v_1, v_2, \ldots, v_N$
$\boldsymbol{\omega}:$	$\omega_1, \omega_2, \ldots, \omega_N$
$\boldsymbol{\delta}:$	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$
$\Omega:$	

MDS with objects

set of N individuals

$V:$	$v_1, v_2, \ldots, v_N$
$\boldsymbol{\omega}:$	$\omega_1, \omega_2, \ldots, \omega_N$
$\boldsymbol{\delta}:$	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$
$\Omega:$	

MDS with objects

set of N individuals

statistical values

V :	$v_1, v_2, \ldots, v_N$
$\boldsymbol{\omega}$:	$\omega_1, \omega_2, \ldots, \omega_N$
$\boldsymbol{\delta}$:	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$
Ω :	

MDS with objects

set of N individuals

statistical values

dissimilarities

V :	$v_1, v_2, \ldots, v_N$
$\boldsymbol{\omega}$:	$\omega_1, \omega_2, \ldots, \omega_N$
$\boldsymbol{\delta}$:	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$
Ω :	

MDS with objects

set of N individuals

statistical values

dissimilarities

V :	$v_1, v_2, \ldots, v_N$
$\boldsymbol{\omega}$:	$\omega_1, \omega_2, \ldots, \omega_N$
$\boldsymbol{\delta}$:	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$
Ω :	bounded region $\subset \mathbb{R}^n$

MDS with objects. Proportions

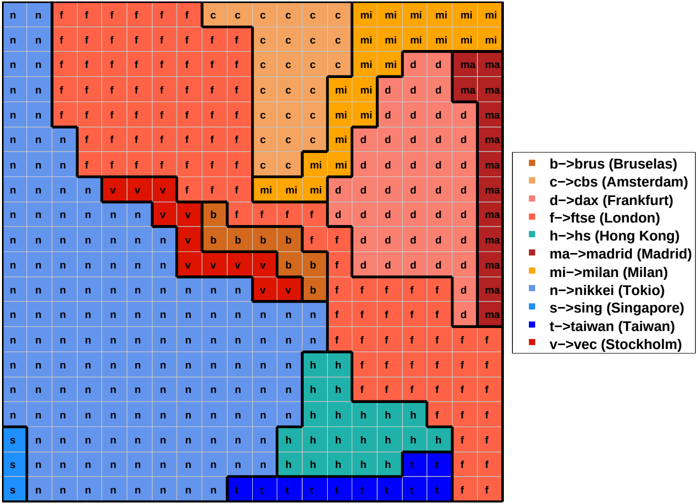
C., Guerrero-Lozano, Romero Morales. Computers & OR, 2017; EJOR, 2018

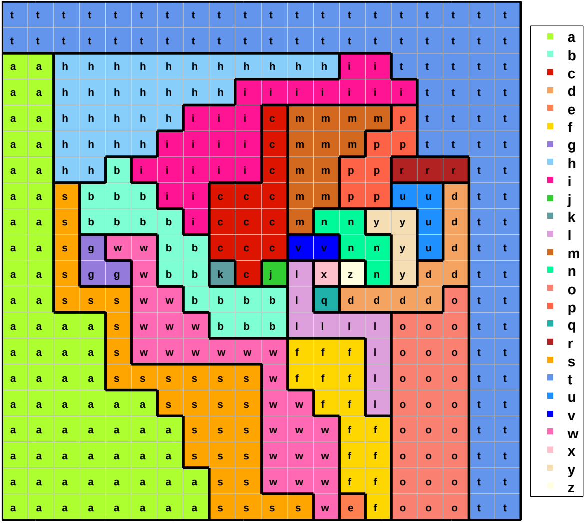
set of N individuals

proportions

dissimilarities

V :	$v_1, v_2, \ldots, v_N$
$\boldsymbol{\omega}$:	$\omega_1, \omega_2, \ldots, \omega_N$
$\boldsymbol{\delta}$:	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$
Ω :	bounded region $\subset \mathbb{R}^n$


n n f f f f f f c c c c c c mi mi mi mi mi mi
n n f f f f f f f f c c c c mi mi mi mi mi mi
n n f f f f f f f f c c c c mi mi d d ma ma
n n f f f f f f f f c c c mi mi d d d ma ma
n n f f f f f f f f c c c mi mi d d d d ma
n n n f f f f f f f c c c mi d d d d d ma
n n n f f f f f f f c c mi mi d d d d d ma
n n n n v v v f f f mi mi mi d d d d d d ma
n n n n n n v v b f f f f d d d d d d ma
n n n n n n n v b b b b f f d d d d d ma
n n n n n n n v v v v b b f d d d d d ma
n n n n n n n n n n v v b f f f f f d ma
n n n n n n n n n n n n n f f f f f d ma
n n n n n n n n n n n n n f f f f f f f
n n n n n n n n n n n n h h f f f f f f
n n n n n n n n n n n n h h f f f f f f
n n n n n n n n n n n n h h h h h f f f
s n n n n n n n n n n h h h h h h h f f
s n n n n n n n n n n h h h h h t t f f
s n n n n n n n n t t t t t t t t t f f
b−>brus (Bruselas)
c−>cbs (Amsterdam)
d−>dax (Frankfurt)
f−>ftse (London)
h−>hs (Hong Kong)
ma−>madrid (Madrid)
mi−>milan (Milan)
n−>nikkei (Tokio)
s−>sing (Singapore)
t−>taiwan (Taiwan)
v−>vec (Stockholm)


t t t t t t t t t t t t t t t t t t t t
t t t t t t t t t t t t t t t t t t t t
a a h h h h h h h h h h h i i t t t t t
a a h h h h h h h i i i i i i i t t t t
a a h h h h h i i i c m m m m p t t t t
a a h h h h i i i i c m m m p p t t t t
a a h h b i i i i i c m m p p r r r t t
a a s b b b i i c c c m m p p u u d t t
a a s b b b b i c c c m n n y y u d t t
a a s g w w b b c c c v v n n y u d t t
a a s g g w b b k c j l x z n y d d t t
a a s s s w w b b b b l q d d d d o t t
a a a a s w w w b b b l l l l o o o t t
a a a a s w w w w w w f f f l o o o t t
a a a a s s s s s s w f f f l o o o t t
a a a a a a s s s s w w f f l o o o t t
a a a a a a a s s s w w w f f o o o t t
a a a a a a a s s s w w w f f o o o t t
a a a a a a a a s s w w w f f o o o t t
a a a a a a a a s s s s w e f o o o t t
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

MDS with objects

set of N individuals

statistical values

dissimilarities

V :	$v_1, v_2, \ldots, v_N$
$\boldsymbol{\omega}$:	$\omega_1, \omega_2, \ldots, \omega_N$
$\boldsymbol{\delta}$:	$\begin{pmatrix} 0 & \delta_{12} & \cdots & \delta_{1N} \\ \delta_{21} & 0 & \cdots & \delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & 0 \end{pmatrix}$
Ω :	bounded region $\subset \mathbb{R}^n$

Modeling distance fit

C., Guerrero, Romero Morales; Mathematical Programming, 2018

- Distance function d, defined on pairs of compact convex sets of $\mathbb{R}^n$, satisfying for any A_1, A_2
  (i) $d \geq 0$ and d is symmetric
  (ii) $d(A_1, A_2) = d(A_1 + z, A_2 + z), \forall z \in \mathbb{R}^n$
  (iii) The function $d_z : z \in \mathbb{R}^n \longmapsto d(z + A_1, A_2)$ is convex and satisfies for all $\theta > 0$ that $d_z(\theta A_1, \theta A_2) = \theta d_{\frac{1}{\theta}z}(A_1, A_2)$.
- Possible choices of d :
  1. The infimum distance, $d(A_1, A_2) = \inf\{\|a_1 - a_2\| : a_1 \in A_1, a_2 \in A_2\}$
  2. The supremum distance, $d(A_1, A_2) = \sup\{\|a_1 - a_2\| : a_1 \in A_1, a_2 \in A_2\}$
  3. The average distance, $d(A_1, A_2) = \dfrac{1}{vol(A_1)vol(A_2)} \displaystyle\int \|a_1 - a_2\| d\mu_1 d\mu_2$,

  where $vol(\cdot)$ denotes the volume of a set and μ_1, μ_2 are probability distributions with support A_1 and A_2.

MDS with objects: objectives

Biobjective optimization problem:

- Distances between objects resemble dissimilarities
- Objects are spread out in Ω

MDS with objects: objectives

Biobjective optimization problem:

- Distances between objects resemble dissimilarities
- Objects are spread out in Ω

Distances resemble dissimilarities

$$
\begin{array}{rlcl}
F_1 : & \mathbb{R}^n \times \ldots \times \mathbb{R}^n \times \mathbb{R}^+ \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\
& (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \tau, \kappa) & \longmapsto & \displaystyle\sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} \left[d(\boldsymbol{c}_i + \tau r_i \mathcal{B}, \boldsymbol{c}_j + \tau r_j \mathcal{B}) - \kappa \delta_{ij} \right]^2 .
\end{array}
$$

MDS with objects: objectives

Biobjective optimization problem:

- Distances between objects resemble dissimilarities
- Objects are spread out in Ω

Distances resemble dissimilarities

$$
\begin{array}{rlcl}
F_1: & \mathbb{R}^n \times \ldots \times \mathbb{R}^n \times \mathbb{R}^+ \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\
& (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \tau, \kappa) & \longmapsto & \displaystyle\sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} \left[d(\boldsymbol{c}_i + \tau r_i \mathcal{B}, \boldsymbol{c}_j + \tau r_j \mathcal{B}) - \kappa \delta_{ij} \right]^2 .
\end{array}
$$

Spread: separate the objects

$$
\begin{array}{rlcl}
F_2: & \mathbb{R}^n \times \ldots \times \mathbb{R}^n \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\
& (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \tau) & \longmapsto & \displaystyle - \sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} d^2(\boldsymbol{c}_i + \tau r_i \mathcal{B}, \boldsymbol{c}_j + \tau r_j \mathcal{B}).
\end{array}
$$

MDS with objects: objectives

Biobjective optimization problem:

- Distances between objects resemble dissimilarities
- Objects are spread out in Ω

Distances resemble dissimilarities

$$
\begin{array}{rlcl}
F_1 : & \mathbb{R}^n \times \ldots \times \mathbb{R}^n \times \mathbb{R}^+ \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\
& (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \tau, \kappa) & \longmapsto & \displaystyle\sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} \left[d(\boldsymbol{c}_i + \tau r_i \mathcal{B}, \boldsymbol{c}_j + \tau r_j \mathcal{B}) - \kappa \delta_{ij} \right]^2 .
\end{array}
$$

Spread: reduce the penetration depth

$$
\begin{array}{rlcl}
F_2^{\Pi} : & \mathbb{R}^n \times \ldots \times \mathbb{R}^n \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\
& (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \tau) & \longmapsto & \displaystyle\sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} \pi^2 \left(\boldsymbol{c}_i + \tau r_i \mathcal{B}, \boldsymbol{c}_j + \tau r_j \mathcal{B} \right) .
\end{array}
$$

MDS with objects: objectives

Biobjective optimization problem:

- Distances between objects resemble dissimilarities
- Objects are spread out in Ω

Distances resemble dissimilarities

$$
\begin{array}{rlcl}
F_1: & \mathbb{R}^n \times \ldots \times \mathbb{R}^n \times \mathbb{R}^+ \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\
 & (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \tau, \kappa) & \longmapsto & \displaystyle\sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} \left[d(\boldsymbol{c}_i + \tau r_i \mathcal{B}, \boldsymbol{c}_j + \tau r_j \mathcal{B}) - \kappa \delta_{ij} \right]^2 .
\end{array}
$$

Spread: separate the centers

$$
\begin{array}{rlcl}
F_2^{\boldsymbol{c}}: & \mathbb{R}^n \times \ldots \times \mathbb{R}^n & \longrightarrow & \mathbb{R}^+ \\
 & (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N) & \longmapsto & -\displaystyle\sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} \left\| \boldsymbol{c}_i - \boldsymbol{c}_j \right\|^2 .
\end{array}
$$

MDS with objects

$$\begin{array}{lll} \min\limits_{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_N,\tau,\kappa} & \lambda F_1(\boldsymbol{c}_1,\ldots,\boldsymbol{c}_N,\tau,\kappa) + (1-\lambda)F_2^*(\boldsymbol{c}_1,\ldots,\boldsymbol{c}_N,\tau) & \\ \text{s.t.} & \boldsymbol{c}_i + \tau r_i \mathcal{B} \subseteq \Omega,\ i = 1,\ldots,N & (VM)^* \\ & \tau \in T & \\ & \kappa \in K, & \end{array}$$

where $K, T \subset \mathbb{R}^+$, $\lambda \in [0,1]$, and F_2^* is either F_2, F_2^Π or $F_2^{\boldsymbol{c}}$ (yielding problems (VM), $(VM)^\Pi$ or $(VM)^{\boldsymbol{c}}$, respectively).

MDS with objects: theoretical results

Proposition

Given $\lambda \in [0,1]$, one has:

- $\lambda F_1 + (1-\lambda) F_2$ is DC,
- $\lambda F_1 + (1-\lambda) F_2^{\Pi}$ is DC,
- $\lambda F_1 + (1-\lambda) F_2^{\boldsymbol{c}}$ is DC.

MDS with objects: theoretical results

Proposition

Given $\lambda \in [0, 1]$, one has:

- $\lambda F_1 + (1 - \lambda)F_2$ is DC,
- $\lambda F_1 + (1 - \lambda)F_2^{\Pi}$ is DC,
- $\lambda F_1 + (1 - \lambda)F_2^{\boldsymbol{c}}$ is DC.

Proposition

The function $\lambda F_1 + (1 - \lambda)F_2$, where d is the infimum distance, can be expressed as a DC function, $\lambda F_1 + (1 - \lambda)F_2 = u - (u - \lambda F_1 + (1 - \lambda)F_2)$, where the quadratic separable convex function u is given by

$$u = \max\{3\lambda - 1, 0\} \cdot \left[\sum_{i=1,\dots,N} 8(N-1)\|\boldsymbol{c}_i\|^2 + \tau^2 \sum_{\substack{i,j=1,\dots,N \\ i \neq j}} \beta_{ij} \right] + 2\lambda\kappa^2 \sum_{\substack{i,j=1,\dots,N \\ i \neq j}} \delta_{ij}^2,$$

where β_{ij} satisfies $\beta_{ij} \geq 2\|r_i b_i - r_j b_j\|^2$ for all $b_i, b_j \in \mathcal{B}$.

DCA to solve (VM)

Optimization problems to solve have the form:

$$\begin{array}{ll} \min\limits_{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_N,\tau,\kappa} & \left\{ \sum\limits_{i=1,\ldots,N} M_i^{\boldsymbol{c}} \|\boldsymbol{c}_i\|^2 + M^\kappa \kappa^2 + M^\tau \tau^2 + \sum\limits_{i=1,\ldots,N} \boldsymbol{c}_i{}^\top \boldsymbol{q}_i^{\boldsymbol{c}} + p^\kappa \kappa + p^\tau \tau \right\} \\ \text{s.t.} & \quad \boldsymbol{c}_i + \tau r_i \mathcal{B} \subseteq \Omega,\ i = 1,\ldots,N \\ & \quad \tau \in T \\ & \quad \kappa \in K, \end{array}$$

for scalars $M_i^{\boldsymbol{c}}$, M^κ, $M^\tau \geq 0$, vectors $\boldsymbol{q}_i^{\boldsymbol{c}}$ and scalars p^κ and p^τ.

DCA to solve (VM)

Optimization problems to solve have the form:

$$\min_{\kappa \in K} \quad \left\{ M^{\kappa}\kappa^2 + p^{\kappa}\kappa \right\} + \min_{\substack{\boldsymbol{c}_i + \tau r_i \mathcal{B} \subseteq \Omega \\ \tau \in T}} \left\{ \sum_{i=1,\ldots,N} M_i^{\boldsymbol{c}_i} \|\boldsymbol{c}_i\|^2 + {\boldsymbol{c}_i}^{\top} \boldsymbol{q}_i^{\boldsymbol{c}} + M^{\tau}\tau^2 + p^{\tau}\tau \right\}$$

for scalars $M_i^{\boldsymbol{c}}$, M^{κ}, $M^{\tau} \geq 0$, vectors $\boldsymbol{q}_i^{\boldsymbol{c}}$ and scalars p^{κ} and p^{τ}.

DCA to solve (VM)

Optimization problems to solve have the form:

$$\min_{\kappa \in K} \left\{ M^\kappa \kappa^2 + p^\kappa \kappa \right\} + \min_{\substack{\boldsymbol{c}_i + \tau r_i \mathcal{B} \subseteq \Omega \\ \tau \in T}} \left\{ \sum_{i=1,\ldots,N} M_i^{\boldsymbol{c}_i} \|\boldsymbol{c}_i\|^2 + \boldsymbol{c}_i^\top \boldsymbol{q}_i^{\boldsymbol{c}} + M^\tau \tau^2 + p^\tau \tau \right\}$$

for scalars $M_i^{\boldsymbol{c}}$, M^κ, $M^\tau \geq 0$, vectors $\boldsymbol{q}_i^{\boldsymbol{c}}$ and scalars p^κ and p^τ

Convex problem in one variable.

Separable in the variables $\boldsymbol{c}_i$ if τ is fixed.

DCA to solve (VM)

Optimization problems to solve have the form:

$$\min_{\kappa\in K} \left\{M^\kappa \kappa^2 + p^\kappa \kappa\right\} + \min_{\substack{\boldsymbol{c}_i+\tau r_i\mathcal{B}\subseteq\Omega\\ \tau\in T}} \left\{\sum_{i=1,\ldots,N} M_i^{\boldsymbol{c}_i}\|\boldsymbol{c}_i\|^2 + \boldsymbol{c}_i{}^\top \boldsymbol{q}_i^{\boldsymbol{c}} + M^\tau\tau^2 + p^\tau\tau\right\}$$

for scalars $M_i^{\boldsymbol{c}}$, M^κ, $M^\tau \geq 0$, vectors $\boldsymbol{q}_i^{\boldsymbol{c}}$ and scalars p^κ and p^τ

Convex problem in one variable.

Alternating strategy:

- Optimization of τ for $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_N$ fixed.
- For a fixed τ, optimize $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_N$ by solving N optimization problems

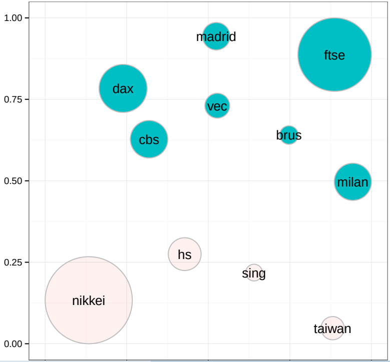
$$\begin{array}{ll}\min\limits_{\boldsymbol{c}_i} & \left\{M_i^{\boldsymbol{c}_i}\|\boldsymbol{c}_i\|^2 + \boldsymbol{c}_i{}^\top \boldsymbol{q}_i^{\boldsymbol{c}}\right\}\\ \text{s.t.} & \boldsymbol{c}_i \in \Omega - \tau r_i\mathcal{B}.\end{array}$$

MDS with objects: Experiments

- Algorithm coded in C on a Windows 8.1 PC Intel® Core™ i7-4500U, 16GB of RAM.
- Quadratic integer programs solved with CPLEX 12.6.
- 3 steps of the alternating algorithm, where each step executes 50 iterations of DCA.
- 100 runs of the multistart strategy, where initial values for $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N$ are uniformly generated in Ω.
- $\lambda = 0.9$.

Visualizing financial markets

V: 11 financial markets across Europe and Asia;
ω: importance regarding to the world market portfolio, Flavin, Hurley and Rousseau, 2002;
$\boldsymbol{\delta}$: correlation between markets, Borg and Groenen, 2005;
$\mathcal{B}$: disc centered at the origin with radius equal to one;
$\Omega = [0,1]^2$.

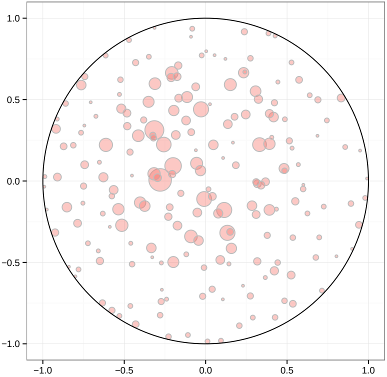


Visualizing financial markets

V: 11 financial markets across Europe and Asia;
ω: importance regarding to the world market portfolio, Flavin, Hurley and Rousseau, 2002;
$\boldsymbol{\delta}$: correlation between markets, Borg and Groenen, 2005;
$\mathcal{B}_1 = [-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}] \times [-\frac{\sqrt{2}}{4}, \frac{\sqrt{2}}{4}]$, $\mathcal{B}_2 = [-\frac{\sqrt{2}}{4}, \frac{\sqrt{2}}{4}] \times [-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}]$;
$\Omega = [0, 1]^2$.

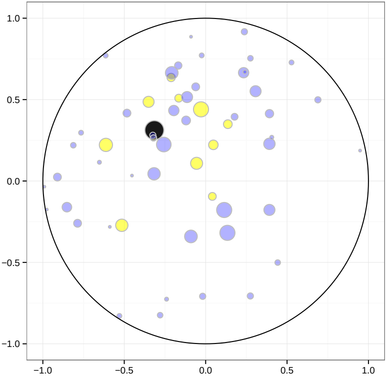
Visualizing a social network

V: 200 musicians;
ω: degree of influence, Dörk, Carpendale and Williamson, 2012;
$\boldsymbol{\delta}$: shortest path in the network;
$\mathcal{B}$: disc centered at the origin with radius equal to one;
$\Omega =$ disc centered at the origin with radius equal to one.

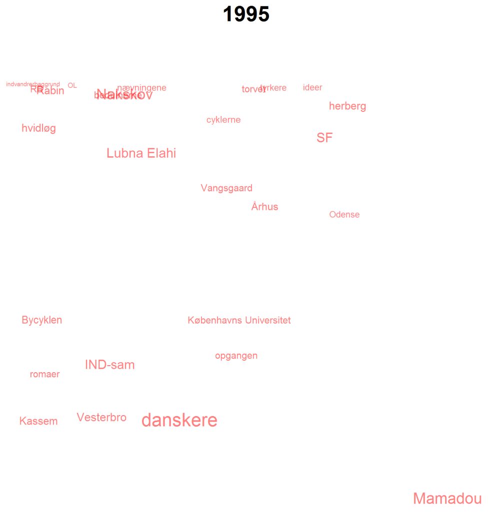


Visualizing a social network

V: 200 musicians;
ω: degree of influence, Dörk, Carpendale and Williamson, 2012;
$\boldsymbol{\delta}$: shortest path in the network;
$\mathcal{B}$: disc centered at the origin with radius equal to one;
$\Omega =$ disc centered at the origin with radius equal to one.



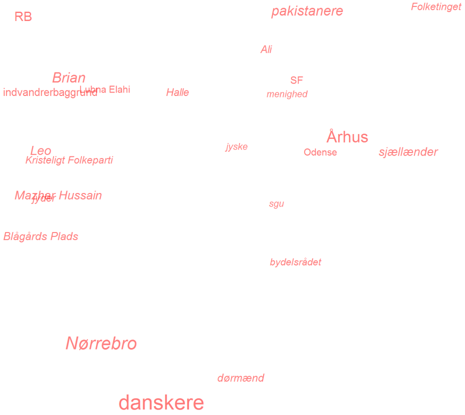
C., Guerrero, Romero Morales. Omega, 2019
C., Guerrero, Hardt, Romero Morales, Big Data, 2018



1995


indvandrerbaggrund
Rabin
OL
Nakskov
torvet
tyrkere
ideer
herberg
cyklerne
hvidløg
SF
Lubna Elahi
Vangsgaard
Århus
Odense
Bycyklen
Københavns Universitet
opgangen
IND-sam
romaer
Kassem
Vesterbro
danskere
Mamadou

1996


RB
pakistanere
Folketinget
Ali
Brian
indvandrerbaggrund
Lubna Elahi
Halle
SF
menighed
Århus
Leo
Kristeligt Folkeparti
jyske
Odense
sjællænder
Mazher Hussain
fynder
sgu
Blågårds Plads
bydelsrådet
Nørrebro
dørmænd
danskere

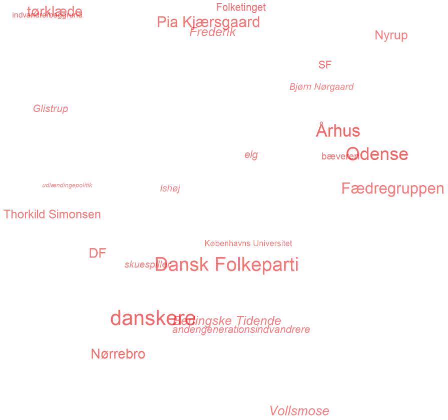
1997

Lubna Elahi
fremmede
Folketinget
Pia Kjærsgaard
Birte Weiss
RB
taxa
indvandrerbaggrund
Fremskridtspartiet
Thorvaldsen
SF
somaliere
Ekstra Bladet
Århus
Odense
danskundervisning
Ishøj
Thorkild Simonsen
Blågårds Plads
familiesammenførte
Dansk Folkeparti
Khader
andengenerationsindvandrere
Nørrebro
urfuglen
danskere
Venstre

1998


indvandrerbaggrund
RB
Folketinget
fremmede
vildsvin
Pia Kjærsgaard
SF
Århus
Nyrup
Odense
Mogens Camre
Fædregruppen
bæveren
Thorkild Simonsen
Cirkeline
Københavns Universitet
integrationsminister
DF
Dansk Folkeparti
Anne Knudsen
Khader
danskere
filmen
Nørrebro

1999


tørklæde
indvandrerbaggrund
Folketinget
Pia Kjærsgaard
Frederik
Nyrup
SF
Bjørn Nørgaard
Glistrup
Århus
elg
bæveren
Odense
udlændingepolitik
Ishøj
Fædregruppen
Thorkild Simonsen
Københavns Universitet
DF
skuespiller
Dansk Folkeparti
danskere
Berlingske Tidende
andengenerationsindvandrere
Nørrebro
Vollsmose

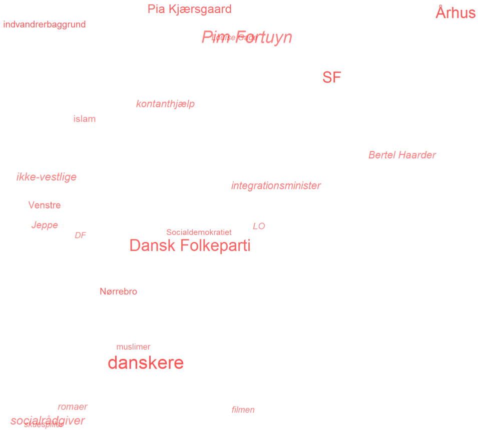
2000

tørklæde
Bispehaven
indvandrerbaggrund
Esbjerg
Pia Kjærsgaard
SF
Folketinget
Nyrup
Århus
Odense
Socialdemokratiet
Københavns Universitet
Venstre
Dansk Folkeparti
Karen Jespersen
Imstr
Berlingske Tidende
arj
andengenerationsindvandrere
Khader
Nørrebro
halalhippier
danskere

2001


indvandrerbaggrund
Pia Kjærsgaard
SF
stationen
Folketinget
Nyrup
islam
Århus
Odense
Ishøj
Københavns Universitet
Dansk Folkeparti
Socialdemokratiet
tvangsægteskaber
Venstre
Karen Jespersen
Berlingske Tidende
Khader
Nørrebro
ECRI
danskere
muslimer
Fogh

2002


Pia Kjærsgaard
Århus
indvandrerbaggrund
Pim Fortuyn
Louise Gade
SF
kontanthjælp
islam
Bertel Haarder
ikke-vestlige
integrationsminister
Venstre
Jeppe
DF
Socialdemokratiet
LO
Dansk Folkeparti
Nørrebro
muslimer
danskere
romaer
filmen
socialrådgiver
skuespiller

2003

Ulla Dahlerup
Pia Kjærsgaard
Folketinget
indvandrerbaggrund
Frederik
århusianske
SF
tørklæde
tosprogede
kontanthjælp
Århus
Odense
Bertel Haarder
kronik
integrationsminister
Dansk Folkeparti
tvangsægteskaber
Nørrebro
Khader
Berlingske Tidende
Khaled Ramadan
POEM
danskere
Fogh

2004

tosprogede
Gogh
indvandrerbaggrund
OL
Folketinget
tørklæde
Statistik
SF
imamer
kontanthjælp
Århus
Odense
islam
moskeer
Kurt
Dansk Folkeparti
Universitet
integrationsminister
tvangsægteskaber
Theo
muslimer
Leyla
Nørrebro
danskere
Hasan

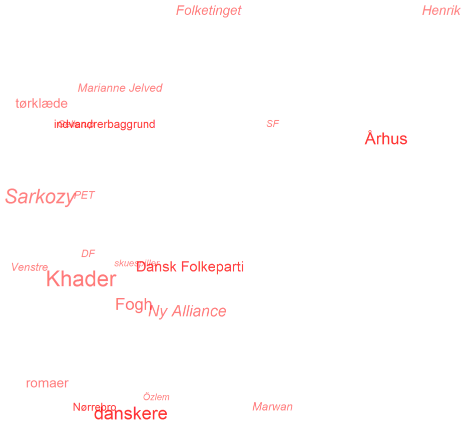
2005

indvandrerbaggrund
kontanthjælp
tosprogede
Folketinget
Pia Kjærsgaard
imamer
ghettoer
SF
Århus
Venstre
Berlingske Tidende
DF
Dansk Folkeparti
skuespiller
Københavns Universitet
Hüseyin Arac
Fogh
muslimer
Nørrebro
Wallait Khan
danskere

2006


kontanthjælp
Muhammed
Pia Kjærsgaard
profeten
tørklæde
mentor
indvandrerbaggrund
imamer
islam
Århus
tegninger
integrationsminister
moskeer
Khader
Nørrebro
Dansk Folkeparti
muslimer
Abu Laban
Fogh
danskere
romaer
Berlingske Tidende

2007


Folketinget
Henrik
Marianne Jelved
tørklæde
indvandrerbaggrund
SF
Århus
Sarkozy
PET
DF
Venstre
skuespiller
Dansk Folkeparti
Khader
Fogh
Ny Alliance
romaer
Özlem
Nørrebro
danskere
Marwan

2008

Mc Cain

Pia Kjærsgaard Folketinget

Hells Angels

Århus

tørklæde

Frederik

vampyr

SF

indvandrerbaggrund

islam

Sarkozy

Obama

DF

Dansk Folkeparti

danskere

Berlingske Tidende

Carlos

Nørrebro

muslimer

Villy Søvndal

Fogh

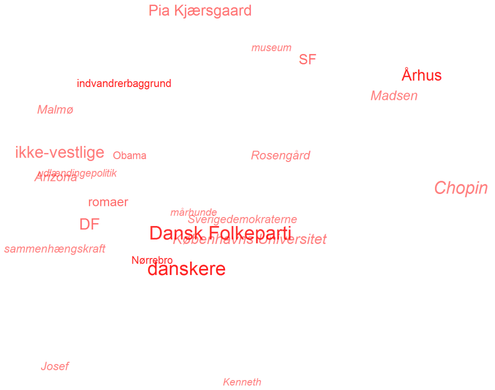
romaer

drenge

2009


indvandrerbaggrund
Hells Angels
bandekrigen
Pia Kjærsgaard
Hans Lassen
tørklæde
SF
skyderier
Århus
Jønke
ikke-vestlige
PET
Tingbjerg
Dansk Folkeparti
DF
Khader
uighurer
romaer
danskere
Mustafa
Nørrebro
rockere
Obama

2010


Pia Kjærsgaard
museum
SF
Århus
indvandrerbaggrund
Madsen
Malmø
ikke-vestlige
Obama
Rosengård
udlændingepolitik
Arizona
Chopin
romaer
mårhunde
Sverigedemokraterne
DF
Dansk Folkeparti
Københavns Universitet
sammenhængskraft
Nørrebro
danskere
Josef
Kenneth

2011


MFP
Hells Angels
Danmarks Statistik
Søren Pind
tosprogede
Folketinget
DSK
blog
indvandrerbaggrund
SF
kronik
Aarhus
PET
udlændingepolitik
rockere
Strauss-Kahn
ikke-vestlige
DF
Dansk Folkeparti
Jamal
stuepige
drabsforsøg
danskere
ORG
Nørrebro

2012


Sleiman
Kristian Thulesen Dahl
Pia Kjærsgaard
Reich
Romney
Wilders
indvandrerbaggrund
Aarhus
Obama
ikke-vestlige
Sarkozy
DF
Dansk Folkeparti
filmen
Eske
flygel
Nørrebro
dingoen
Mir
skuespiller
danskere
Hamid Rahmati

2013


Krasnik
Republikanerne
Fremskridtspartiet
tørklæde
SF
indvandrerbaggrund
islam
Nyrup
FatmaØktem
Gyldent Daggry
Aarhus
ph
Dansk Folkeparti
DF
romaer
Khader
muslimer
Københavns Universitet
Lars Hedegaard
Yahya Hassan
digte
danskere
ulve

2014


Krim
kontanthjælp
indvandrerbaggrund
SF
velfærdsturisme
elg
Grønland
PET
Aarhus
ikke-vestlige
Obama
Frank Jensen
bogen
romaer
Dansk Folkeparti
Manu Sareen
DF
skuespiller
Yahya Hassan
digte
danskere
Nørrebro
ulve
Basim

2015


Charlie Hebdo
Facebook
Egtvedpigen
Islamisk Stat
Inger Støjberg
Pia Kjærsgaard
Fremskridtspartiet
Folketinget
Kristian Thulesen Dahl
Fryd
islam
indvandrerbaggrund
Lars Løkke Rasmussen
kronik
Henrik
Venstre
Manu Sareen
DF
Khader
Dansk Folkeparti
danskere
Majid
Nørrebro
Dieudonné

Supervised classification

Supervised classification. The framework

- Given: set I of individuals, each $i \in I$ with associated

Supervised classification. The framework

- Given: set I of individuals, each $i \in I$ with associated
  - A vector $\mathbf{x}_i \in \mathcal{X}$, assumed here $\mathcal{X} \subset \mathbb{R}^m$
  - A label y_i, assumed here to be in $\{-1, +1\}$
- Seen as a sample from $(\mathbf{X}, Y)$, with unknown distribution

Supervised classification. The framework

- Given: set I of individuals, each $i \in I$ with associated
  - A vector $\mathbf{x}_i \in \mathcal{X}$, assumed here $\mathcal{X} \subset \mathbb{R}^m$
  - A label y_i, assumed here to be in $\{-1, +1\}$
- Seen as a sample from $(\mathbf{X}, Y)$, with unknown distribution
- Goal: to infer from I a classifier $\varphi : \mathcal{X} \longrightarrow \{-1, 1\}$ so that we can classify any object just by knowing $\mathbf{x}$

Supervised classification. The framework

- Given: set I of individuals, each $i \in I$ with associated
  - A vector $\mathbf{x}_i \in \mathcal{X}$, assumed here $\mathcal{X} \subset \mathbb{R}^m$
  - A label y_i, assumed here to be in $\{-1, +1\}$
- Seen as a sample from $(\mathbf{X}, Y)$, with unknown distribution
- Goal: to infer from I a classifier $\varphi : \mathcal{X} \longrightarrow \{-1, 1\}$ so that we can classify any object just by knowing $\mathbf{x}$
- Linear classifier $\varphi : \mathbf{x} = (x_1, \ldots, x_m) \longmapsto \{-1, 1\}$:
  - score function:

$$\mathbf{x} = (x_1, \ldots, x_m) \longmapsto \omega_1 x_1 + \ldots + \omega_n x_m + \beta$$

  - $\varphi(x) = \begin{cases} 1, & \text{if } \omega_1 x_1 + \ldots + \omega_n x_m + \beta > 0 \\ -1, & \text{else} \end{cases}$

Supervised classification. The framework

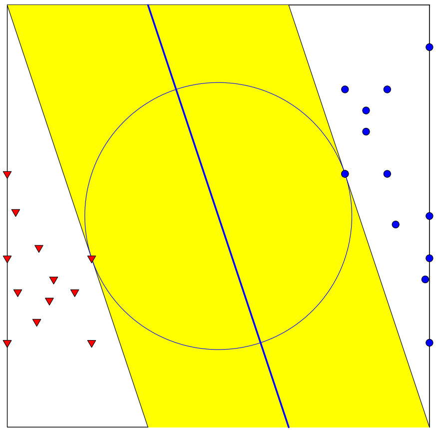
- Given: set I of individuals, each $i \in I$ with associated
  - A vector $\mathbf{x}_i \in \mathcal{X}$, assumed here $\mathcal{X} \subset \mathbb{R}^m$
  - A label y_i, assumed here to be in $\{-1, +1\}$
- Seen as a sample from $(\mathbf{X}, Y)$, with unknown distribution
- Goal: to infer from I a classifier $\varphi : \mathcal{X} \longrightarrow \{-1, 1\}$ so that we can classify any object just by knowing $\mathbf{x}$
- Linear classifier $\varphi : \mathbf{x} = (x_1, \ldots, x_m) \longmapsto \{-1, 1\}$:
  - score function:

$$\mathbf{x} = (x_1, \ldots, x_m) \longmapsto \omega_1 x_1 + \ldots + \omega_n x_m + \beta$$

  - $\varphi(x) = \begin{cases} 1, & \text{if } \omega_1 x_1 + \ldots + \omega_n x_m + \beta > 0 \\ -1, & \text{else} \end{cases}$
- Problem: how to infer from I the coefficients $\omega = (\omega_1, \ldots, \omega_n), \beta$?

SVM

- Roughly speaking, SVM finds the hyperplane $\omega_1 x_1 + \ldots + \omega_m x_m + \beta = 0$ separating most the sets $\{\mathbf{x}_i : i \in I, y_i = 1\}$ and $\{\mathbf{x}_i : i \in I, y_i = -1\}$



SVM

Convex quadratic optimization problem with linear constraints:

C., and Romero Morales, "Supervised classification and mathematical optimization", *Computers & Operations Research*, 2013.

Duarte Silva, "Optimization approaches to supervised classification", *EJOR*, 2017.

SVM

Convex quadratic optimization problem with linear constraints:

$$\begin{array}{lll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C \sum_{i \in I} \xi_i & \\ \text{s.t.} & y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\ & \xi_i \geq 0 & i \in I \end{array}$$

C., and Romero Morales, "Supervised classification and mathematical optimization", *Computers & Operations Research*, 2013.

Duarte Silva, "Optimization approaches to supervised classification", *EJOR*, 2017.

SVM

Convex quadratic optimization problem with linear constraints:

$$\begin{array}{lll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & \\ \text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\ & \xi_i \geq 0 & i \in I \end{array}$$

$$\begin{array}{lll} \max_{\lambda} & \sum_{i\in I}\lambda_i - \frac{1}{2}\sum_{ij}\lambda_i y_i \lambda_j y_j \mathbf{x}_i^\top \mathbf{x}_j & \\ \text{s.t.} & \sum_{i\in I}\lambda_i y_i = 0 & \\ & 0 \leq \lambda_i \leq \frac{C}{2} & i \in I \end{array}$$

C., and Romero Morales, "Supervised classification and mathematical optimization", *Computers & Operations Research*, 2013.

Duarte Silva, "Optimization approaches to supervised classification", *EJOR*, 2017.

SVM

Convex quadratic optimization problem with linear constraints:

$$
\begin{array}{lll}
\min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & \\
\text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
 & \xi_i \geq 0 & i \in I
\end{array}
$$

$$
\begin{array}{lll}
\max_{\lambda} & \sum_{i\in I}\lambda_i - \frac{1}{2}\sum_{ij}\lambda_i y_i \lambda_j y_j K(\mathbf{x}_i, \mathbf{x}_j) & \\
\text{s.t.} & \sum_{i\in I}\lambda_i y_i = 0 & \\
 & 0 \leq \lambda_i \leq \frac{C}{2} & i \in I
\end{array}
$$

C., and Romero Morales, "Supervised classification and mathematical optimization", *Computers & Operations Research*, 2013.

Duarte Silva, "Optimization approaches to supervised classification", *EJOR*, 2017.

Kernels

$K(\mathbf{x}_i, \mathbf{x}_j) = \ldots$

- $\mathbf{x}_i^\top \mathbf{x}_j$ (linear kernel)
- $(1 + \mathbf{x}_i^\top \mathbf{x}_j)^d$ (polynomial kernel)
- $e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}$ (gaussian kernel)
- $\sum_k \theta_k e^{-\gamma_k \|\mathbf{x}_i - \mathbf{x}_j\|^2}$
- ... many more (not only for $\mathbf{x}$ in a dot product space)

Cristianini and Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*, 2000.

Hofmann, Schölkopf and Smola, "Kernel methods in Machine Learning", *Annals of Statistics*, 2008.

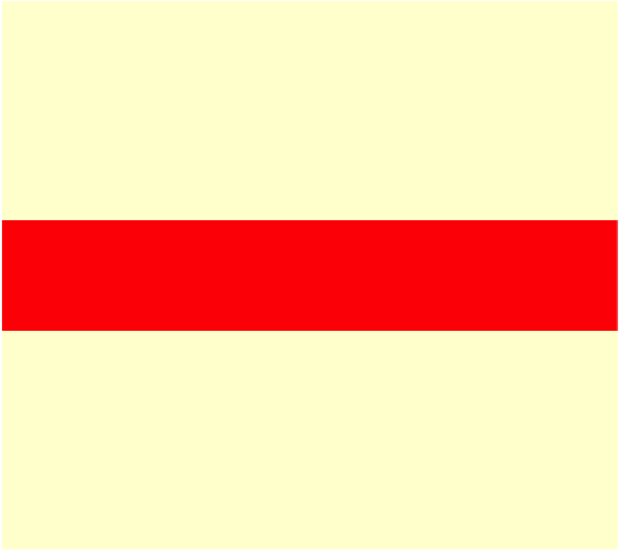
Parameters tuning

$$\begin{array}{lll} \max_\lambda & \sum_{i\in I}\lambda_i - \frac{1}{2}\sum_{ij}\lambda_i y_i \lambda_j y_j e^{-\gamma\|\mathbf{x}_i-\mathbf{x}_j\|^2} & \\ \text{s.t.} & \sum_{i\in I}\lambda_i y_i = 0 & \\ & 0 \le \lambda_i \le \frac{C}{2} & i \in I \end{array} \qquad (P_{I,C,\gamma})$$

C, γ: k-fold crossvalidation

- I : split in k blocks of similar size, $I_1, \ldots, I_k$
- for each pair C, γ in a grid (e.g. $2^{-12} \ldots, 2^{12}$), estimate $acc(C, \gamma)$:
  - for each $i = 1, \ldots, k$
    - solve $(P_{I \setminus I_i, C, \gamma})$, yielding λ^i, β (via KKT)
    - calculate $acc(C, \gamma, I_i)$, fraction of correctly classified in I_i if classifier with λ^i, β were used
  - $acc(C, \gamma) = \frac{1}{k}\sum_{i=1}^{k} acc(C, \gamma, I_i)$

Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection", *IJCAI*, 1995.



The performance measure

$$
\begin{array}{lll}
\min_{\omega,\beta,\xi} & \|\omega\|^2 + C \sum_{i\in I} \xi_i & \\
\text{s.t.} & y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
 & \xi_i \geq 0 & i \in I
\end{array}
$$

The performance measure

$$\begin{array}{llll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C \sum_{i \in I} \xi_i & & \\ \text{s.t.} & y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I & \qquad \omega, \beta, C \\ & \xi_i \geq 0 & i \in I & \end{array}$$

The performance measure

$$\begin{array}{llll}
\min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & & \\
\text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I & \\
 & \xi_i \geq 0 & i \in I &
\end{array} \qquad \omega, \beta, C$$

The performance measure

- $\{(\mathbf{x}_i, y_i) : i \in I\}$: seen as a random sample of $(\mathbf{X}, Y)$
- Accuracy: $acc = P(Y(\omega^\top \mathbf{X} + \beta) > 0)$

The performance measure

$$\begin{array}{llll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & & \\ \text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I & \qquad \omega, \beta, C \\ & \xi_i \geq 0 & i \in I & \end{array}$$

The performance measure

- $\{(\mathbf{x}_i, y_i) : i \in I\}$: seen as a random sample of $(\mathbf{X}, Y)$
- Accuracy: $acc = P(Y(\omega^\top \mathbf{X} + \beta) > 0)$
- Distribution of $(\mathbf{X}, Y)$: Unknown

(Asymmetric) costs

- Benítez-Peña, Blanquero, C., Ramírez-Cobo, "On Support Vector Machines under a multiple-cost scenario". Advances in Data Analysis and Classification, 2017.
- C., Martín-Barragán, Romero Morales. "Multi-group support vector machines with measurement costs: A biobjective approach". Discrete Applied Mathematics, 2008.
- He, Ma. *Imbalanced learning: foundations, algorithms, and applications.* Wiley, 2013.
- Maldonado, Pérez, Bravo. "Cost-based feature selection for support vector machines: An application in credit scoring". EJOR, 2017.
- Prati, Batista, Duarte Silva. "Class imbalance revisited: a new experimental setup to assess the performance of treatment methods". Knowledge and Information Systems. 2015.
- Turney. "Types of cost in inductive concept learning". 2002.

Performance measures

$$
\begin{array}{lll}
\min_{\omega,\beta,\xi} & \|\omega\|^2 + C \sum_{i \in I} \xi_i & \\
\text{s.t.} & y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
 & \xi_i \geq 0 & i \in I
\end{array}
\qquad \omega, \beta, C
$$

Performance measures $\pi(\omega, \beta)$:

- Accuracy: $acc = P(Y(\omega^\top \mathbf{X} + \beta) > 0)$

Performance measures

$$\begin{array}{llll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & & \\ \text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\ & \xi_i \geq 0 & i \in I \end{array} \qquad \omega, \beta, C$$

Performance measures $\pi(\omega, \beta)$:

- Accuracy: $acc = P(Y(\omega^\top \mathbf{X} + \beta) > 0)$
- Sensitivity: $TPR = P(\omega^\top \mathbf{X} + \beta > 0 | Y = 1)$
- Specificity: $TNR = P(\omega^\top \mathbf{X} + \beta < 0 | Y = -1)$
- Youden's index:
  $J = TPR + TNR - 1 = P(\omega^\top \mathbf{X} + \beta > 0 | Y = 1) + P(\omega^\top \mathbf{X} + \beta < 0 | Y = -1) - 1$
- Positive Predictive Value: $PPV = P(Y = 1 | \omega^\top \mathbf{X} + \beta > 0)$
- Negative Predictive Value: $NPV = P(Y = -1 | \omega^\top \mathbf{X} + \beta < 0)$
- ...

- Performance measures $\pi_\ell(\omega, \beta), \ell \in L$
- Threshold values γ_ℓ for π_ℓ, $\ell \in L$
- I : training sample $\{(\mathbf{x}_i, y_i) : \ i \in I\}$

- Performance measures $\pi_\ell(\omega, \beta), \ell \in L$
- Threshold values γ_ℓ for π_ℓ, $\ell \in L$
- I : training sample $\{(\mathbf{x}_i, y_i) : i \in I\}$
- J : anchor sample $\{(\mathbf{x}_j, y_j) : j \in J\}$ ($J \cap I = \emptyset$)

- Performance measures $\pi_\ell(\omega, \beta), \ell \in L$
- Threshold values γ_ℓ for π_ℓ, $\ell \in L$
- I : training sample $\{(\mathbf{x}_i, y_i) : i \in I\}$
- J : anchor sample $\{(\mathbf{x}_j, y_j) : j \in J\}$ ($J \cap I = \emptyset$)
- Estimates of performance measures: $\widehat{\pi_\ell}(\omega, \beta; J), \ell \in L$

- Performance measures $\pi_\ell(\omega, \beta), \ell \in L$
- Threshold values γ_ℓ for π_ℓ, $\ell \in L$
- I : training sample $\{(\mathbf{x}_i, y_i) : i \in I\}$
- J : anchor sample $\{(\mathbf{x}_j, y_j) : j \in J\}$ ($J \cap I = \emptyset$)
- Estimates of performance measures: $\widehat{\pi_\ell}(\omega, \beta; J), \ell \in L$
- Desired: $\pi_\ell(\omega, \beta) \geq \gamma_\ell$, $\ell \in L$

- Performance measures $\pi_\ell(\omega, \beta), \ell \in L$
- Threshold values γ_ℓ for π_ℓ, $\ell \in L$
- I : training sample $\{(\mathbf{x}_i, y_i) : i \in I\}$
- J : anchor sample $\{(\mathbf{x}_j, y_j) : j \in J\}$ ($J \cap I = \emptyset$)
- Estimates of performance measures: $\widehat{\pi}_\ell(\omega, \beta; J), \ell \in L$
- Desired: $\pi_\ell(\omega, \beta) \geq \gamma_\ell$, $\ell \in L$
- Imposed: $\widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^*$, $\ell \in L$

- Performance measures $\pi_\ell(\omega, \beta), \ell \in L$
- Threshold values γ_ℓ for π_ℓ, $\ell \in L$
- I : training sample $\{(\mathbf{x}_i, y_i) : i \in I\}$
- J : anchor sample $\{(\mathbf{x}_j, y_j) : j \in J\}$ ($J \cap I = \emptyset$)
- Estimates of performance measures: $\widehat{\pi}_\ell(\omega, \beta; J), \ell \in L$
- Desired: $\pi_\ell(\omega, \beta) \geq \gamma_\ell$, $\ell \in L$
- Imposed: $\widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma^*_\ell$, $\ell \in L$

Standard approach

$$\begin{array}{lll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C \sum_{i \in I} \xi_i & \\ \text{s.t.} & y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\ & \xi_i \geq 0 & i \in I \end{array}$$

- Performance measures $\pi_\ell(\omega, \beta), \ell \in L$
- Threshold values γ_ℓ for π_ℓ, $\ell \in L$
- I : training sample $\{(\mathbf{x}_i, y_i) : i \in I\}$
- J : anchor sample $\{(\mathbf{x}_j, y_j) : j \in J\}$ ($J \cap I = \emptyset$)
- Estimates of performance measures: $\widehat{\pi}_\ell(\omega, \beta; J), \ell \in L$
- Desired: $\pi_\ell(\omega, \beta) \geq \gamma_\ell$, $\ell \in L$
- Imposed: $\widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^*$, $\ell \in L$

Constrained approach

$$
\begin{array}{lll}
\min_{\omega,\beta,\xi} & \|\omega\|^2 + C \sum_{i \in I} \xi_i & \\
\text{s.t.} & y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
& \xi_i \geq 0 & i \in I \\
& \widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^* & \ell \in L
\end{array}
$$

Adding constraints to an SVM model

$$
\begin{array}{lll}
\min_{\omega,\beta,\xi} & \|\omega\|^2 + C \sum_{i \in I} \xi_i & \\
\text{s.t.} & y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
 & \xi_i \geq 0 & i \in I \\
 & (\omega, \beta) \in \Omega &
\end{array}
$$

Ω : some (polyhedral) regions forced to be in one side of the separating hyperplane

Adding constraints to an SVM model

$$\begin{array}{lll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & \\ \text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\ & \xi_i \geq 0 & i \in I \\ & (\omega,\beta) \in \Omega & \end{array}$$

Ω : some (polyhedral) regions forced to be in one side of the separating hyperplane

- C., and Plastria, "Optimal expected-distance separating halfspace", *Maths of OR*, 2008.
- Fung, , Mangasarian, and Shavlik, "Knowledge-based support vector machine classifiers". In *Advances in NIPS*, 2002
- Lauer and Bloch, "Incorporating prior knowledge in support vector machines for classification: A review", *Neurocomputing*, 2008.
- Mangasarian, "Knowledge-based linear programming", *SIAM Journal on Optimization*, 2005.
- Mangasarian and Wild, "Nonlinear knowledge-based classification", *IEEE Transactions on Neural Networks*, 2008.

- Desired: $\pi_\ell(\omega, \beta) \geq \gamma_\ell$, $\ell \in L$
- Imposed: $\widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^*$, $\ell \in L$

- Desired: $\pi_\ell(\omega, \beta) \geq \gamma_\ell$, $\ell \in L$
- Imposed: $\widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^*$, $\ell \in L$
- γ_ℓ^* : so that H_0 cannot be rejected in the test hypothesis

$$\begin{cases} H_0: & \pi_\ell(\omega, \beta) \geq \gamma_\ell \\ H_1: & \pi_\ell(\omega, \beta) < \gamma_\ell \end{cases}$$

- Desired: $\pi_\ell(\omega, \beta) \geq \gamma_\ell$, $\ell \in L$
- Imposed: $\widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^*$, $\ell \in L$
- γ_ℓ^* : so that H_0 cannot be rejected in the test hypothesis

$$\begin{cases} H_0 : & \pi_\ell(\omega, \beta) \geq \gamma_\ell \\ H_1 : & \pi_\ell(\omega, \beta) < \gamma_\ell \end{cases}$$

Building γ_ℓ^*

- Hoeffding's inequality: for $Z_1, \ldots, Z_n$ *i.i.d.*, $Be(p)$, $P(\bar{Z} - p \geq c) \leq e^{-2nc^2}$.
- $100(1-\alpha)\%$ CI for p :

$$\left(\bar{Z} - \sqrt{\frac{\log \alpha}{-2n}}, 1\right)$$

- Imposing $p_0 \in CI$ means

$$\bar{Z} \geq p_0 + \sqrt{\frac{\log \alpha}{-2n}}$$

- Desired: $\pi_\ell(\omega, \beta) \geq \gamma_\ell$, $\ell \in L$
- Imposed: $\widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^*$, $\ell \in L$
- γ_ℓ^* : so that H_0 cannot be rejected in the test hypothesis

$$\begin{cases} H_0 : & \pi_\ell(\omega, \beta) \geq \gamma_\ell \\ H_1 : & \pi_\ell(\omega, \beta) < \gamma_\ell \end{cases}$$

Building γ_ℓ^*

- Hoeffding's inequality: for $Z_1, \ldots, Z_n$ $i.i.d.$, $Be(p)$, $P(\bar{Z} - p \geq c) \leq e^{-2nc^2}$.
- $100(1-\alpha)\%$ CI for p :

$$\left(\bar{Z} - \sqrt{\frac{\log \alpha}{-2n}}, 1\right)$$

- Imposing $p_0 \in CI$ means

$$\bar{Z} \geq p_0 + \sqrt{\frac{\log \alpha}{-2n}}$$

- $\gamma_\ell^* = \gamma_\ell + \sqrt{\frac{\log \alpha}{-2|J|}}$

Feasibility?

Always feasible:

$$\begin{array}{lll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C \sum_{i\in I} \xi_i & \\ \text{s.t.} & y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\ & \xi_i \geq 0 & i \in I \end{array}$$

Maybe unfeasible:

$$\begin{array}{lll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C \sum_{i\in I} \xi_i & \\ \text{s.t.} & y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\ & \xi_i \geq 0 & i \in I \\ & \widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^* & \ell \in L \end{array}$$

$$\begin{array}{lll}
\min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & \\
\text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
& \xi_i \geq 0 & i \in I \\
& \widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^* & \ell \in L
\end{array}$$

$$\begin{array}{lll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & \\ \text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\ & \xi_i \geq 0 & i \in I \\ & \widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^* & \ell \in L \end{array} \quad z_j = \begin{cases} 1, & \text{if } y_j\left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 \\ 0, & \text{else} \end{cases} \quad j \in J$$

$$
\begin{array}{lll}
\min_{\omega,\beta,\xi} & \|\omega\|^2 + C \sum_{i \in I} \xi_i & \\
\text{s.t.} & y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
& \xi_i \geq 0 & i \in I \\
& \widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^* & \ell \in L
\end{array}
\qquad
z_j = \begin{cases} 1, & \text{if } y_j \left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 \\ 0, & \text{else} \end{cases} \qquad j \in J
$$

$$\widehat{TPR}(\omega, \beta; J) \geq \gamma$$

$$\sum_{j \in J : y_j = 1} z_j \geq \gamma \, \# \left(\{j \in J : y_j = 1\}\right)$$

$$\begin{array}{lll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & \\ \text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1-\xi_i & i \in I \\ & \xi_i \geq 0 & i \in I \\ & \widehat{\pi}_\ell(\omega,\beta;J) \geq \gamma_\ell^* & \ell \in L \end{array} \qquad z_j = \left\{ \begin{array}{ll} 1, & \text{if } y_j\left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 \\ 0, & \text{else} \end{array} \right. \qquad j \in J$$

$$\widehat{TNR}(\omega,\beta;J) \geq \gamma$$

$$\sum_{j\in J: y_j=-1} z_j \geq \gamma\,\#\left(\{j \in J : y_j = -1\}\right)$$

$$\begin{array}{lll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & \\ \text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\ & \xi_i \geq 0 & i \in I \\ & \widehat{\pi}_\ell(\omega,\beta;J) \geq \gamma_\ell^* & \ell \in L \end{array} \qquad z_j = \left\{ \begin{array}{ll} 1, & \text{if } y_j\left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 \\ 0, & \text{else} \end{array} \right. \qquad j \in J$$

$$\widehat{J}(\omega,\beta;J) \geq \gamma$$

$$\sum_{j\in J} z_j \geq \gamma\,\#\,(J)$$

$$
\begin{array}{lll}
\min_{\omega,\beta,\xi} & \|\omega\|^2 + C \sum_{i\in I} \xi_i & \\
\text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
 & \xi_i \geq 0 & i \in I \\
 & \widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^* & \ell \in L
\end{array}
\qquad
z_j = \begin{cases} 1, & \text{if } y_j\left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 \\ 0, & \text{else} \end{cases} \qquad j \in J
$$

$$\widehat{TPR_m}(\omega, \beta; J) \geq \gamma$$

$$\sum_{j\in J : u_j = m} z_j \geq \gamma \, \# \left(\{j \in J : u_j = m\}\right)$$

$$\begin{array}{lll} \min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & \\ \text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\ & \xi_i \geq 0 & i \in I \\ & \widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^* & \ell \in L \end{array}$$

$$z_j = \left\{ \begin{array}{ll} 1, & \text{if } y_j\left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 \\ 0, & \text{else} \end{array} \right. \quad j \in J$$

$$\widehat{PPV}(\omega, \beta; J) \geq \gamma$$

$$(1-\gamma)\, prev_+ \sum_{j\in J: y_j = 1} z_j - \gamma\,(1 - prev_+) \sum_{j \in J: y_j = -1} (1 - z_j) \geq 0$$

$$
\begin{array}{lll}
\min_{\omega,\beta,\xi} & \|\omega\|^2 + C \sum_{i \in I} \xi_i & \\
\text{s.t.} & y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
 & \xi_i \geq 0 & i \in I \\
 & \widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^* & \ell \in L
\end{array}
\qquad
z_j = \left\{ \begin{array}{ll} 1, & \text{if } y_j \left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 \\ 0, & \text{else} \end{array} \right. \quad j \in J
$$

$$
\widehat{NPV}(\omega, \beta; J) \geq \gamma
$$

$$
(1-\gamma)\, prev_- \sum_{j \in J: y_j = -1} z_j - \gamma \left(1 - prev_-\right) \sum_{j \in J: y_j = 1} \left(1 - z_j\right) \geq 0
$$

$$
\begin{array}{lll}
\min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & \\
\text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
& \xi_i \geq 0 & i \in I \\
& \widehat{\pi}_\ell(\omega,\beta;J) \geq \gamma_\ell^* & \ell \in L
\end{array}
\qquad
z_j = \begin{cases} 1, & \text{if } y_j\left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 \\ 0, & \text{else} \end{cases} \quad j \in J
$$

$$\widehat{\pi}_\ell(\omega,\beta;J) \geq \gamma_\ell^*$$

$$\mathbf{a}_\ell^\top \mathbf{z} \geq b_\ell$$

$$
\begin{array}{lll}
\min_{\omega,\beta,\xi} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & \\
\text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
 & \xi_i \geq 0 & i \in I \\
 & \widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^* & \ell \in L
\end{array}
\qquad
z_j = \begin{cases} 1, & \text{if } y_j\left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 \\ 0, & \text{else} \end{cases} \quad j \in J
$$

$$\widehat{\pi}_\ell(\omega, \beta; J) \geq \gamma_\ell^*$$

$$\mathbf{a}_\ell^\top \mathbf{z} \geq b_\ell$$

$$
\begin{array}{lll}
\min_{\omega,\beta,\xi,\mathbf{z}} & \|\omega\|^2 + C\sum_{i\in I}\xi_i & \\
\text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
 & \xi_i \geq 0 & i \in I \\
 & \mathbf{a}_\ell^\top z \geq b_\ell & \ell \in L \\
 & z_\ell \in \{0, 1\} & \ell \in L \\
 & y_j\left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 - M(1 - z_j) & j \in J
\end{array}
$$

$$
\begin{array}{lll}
\min_{\omega,\beta,\xi,\mathbf{z}} & \|\omega\|^2 + C \sum_{i \in I} \xi_i & \\
\text{s.t.} & y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\
 & \xi_i \geq 0 & i \in I \\
 & \mathbf{a}_l^\top z \geq b_l & l \in L \\
 & z_\ell \in \{0, 1\} & \ell \in L \\
 & y_j \left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 - M(1 - z_j) & j \in J
\end{array}
$$

- Denote $J(z) = \{j \in J : z_j = 1\}$

$$
\begin{array}{llll}
\min_{\mathbf{z}} & & & \min_{\omega,\beta,\xi} \quad \omega^\top \omega + C \sum_{i \in I} \xi_i \\
\text{s.t.} & z_\ell \in \{0,1\} & \ell \in L & \text{s.t.} \quad y_i \left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i \quad i \in I \\
& \mathbf{a}_\ell^\top z \geq b_\ell & l \in L & \qquad\quad y_j \left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 \quad j \in J(\mathbf{z}) \\
& & & \qquad\quad \xi_i \geq 0 \quad i \in I
\end{array}
$$

- Denote $J(z) = \{j \in J : z_j = 1\}$

$$\begin{array}{llll|lll} \min_{\mathbf{z}} & & & \min_{\omega,\beta,\xi} & \omega^\top\omega + C\sum_{i\in I}\xi_i & \\ \text{s.t.} & z_\ell \in \{0,1\} & \ell \in L & \text{s.t.} & y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i & i \in I \\ & \mathbf{a}_\ell^\top z \geq b_\ell & l \in L & & y_j\left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 & j \in J(\mathbf{z}) \\ & & & & \xi_i \geq 0 & i \in I \end{array}$$

KKT conditions for inner problem ($\mathbf{z}$ fixed)

$$\begin{array}{rcll} \omega & = & \sum_{s\in I}\lambda_s y_s \mathbf{x}_s + \sum_{t\in J(\mathbf{z})}\mu_t y_t \mathbf{x}_t & \\ 0 & = & \sum_{s\in I}\lambda_s y_s + \sum_{t\in J(\mathbf{z})}\mu_t y_t & \\ 0 & \leq & \lambda_s \leq C/2 & s \in I \\ 0 & \leq & \mu_t & t \in J(\mathbf{z}) \end{array}$$

- Denote $J(z) = \{j \in J :\, z_j = 1\}$

$$
\begin{array}{llll}
\min_{\mathbf{z}} & & & \min_{\omega,\beta,\xi} \quad \omega^\top\omega + C\sum_{i\in I}\xi_i \\
\text{s.t.} & z_\ell \in \{0,1\} & \ell \in L & \text{s.t.} \quad y_i\left(\omega^\top \mathbf{x}_i + \beta\right) \geq 1 - \xi_i \quad i \in I \\
 & \mathbf{a}_\ell^\top z \geq b_\ell & l \in L & \qquad\;\; y_j\left(\omega^\top \mathbf{x}_j + \beta\right) \geq 1 \qquad j \in J(\mathbf{z}) \\
 & & & \qquad\;\; \xi_i \geq 0 \qquad\qquad\qquad i \in I
\end{array}
$$

KKT conditions for inner problem ($\mathbf{z}$ fixed)

$$
\begin{array}{rcll}
\omega & = & \sum_{s\in I}\lambda_s y_s \mathbf{x}_s + \sum_{t\in J}\mu_t y_t \mathbf{x}_t & \\
0 & = & \sum_{s\in I}\lambda_s y_s + \sum_{t\in J}\mu_t y_t & \\
0 & \leq & \lambda_s \leq C/2 & s \in I \\
0 & \leq & \mu_t \leq M z_t & t \in J
\end{array}
$$

(Partial) Dual

$$
\begin{array}{lll}
\min_{\lambda,\mu,\beta,\xi,\mathbf{z}} & \left(\sum_{s\in I}\lambda_s y_s \mathbf{x}_s + \sum_{t\in J}\mu_t y_t \mathbf{x}_t\right)^\top \left(\sum_{s\in I}\lambda_s y_s \mathbf{x}_s + \sum_{t\in J}\mu_t y_t \mathbf{x}_t\right) & \\
& +C\sum_{i\in I}\xi_i & \\
\text{s.t.} & z_\ell \in \{0,1\} & \ell \in L \\
& \mathbf{a}_\ell^\top z \geq b_\ell & \ell \in L \\
& y_i\left(\left(\sum_{s\in I}\lambda_s y_s \mathbf{x}_s + \sum_{t\in J}\mu_t y_t \mathbf{x}_t\right)^\top \mathbf{x}_i + \beta\right) \geq 1-\xi_i & i \in I \\
& y_j\left(\left(\sum_{s\in I}\lambda_s y_s \mathbf{x}_s + \sum_{t\in J}\mu_t y_t \mathbf{x}_t\right)^\top \mathbf{x}_j + \beta\right) \geq 1-M(1-z_j) & j \in J \\
& \xi_i \geq 0 & i \in I \\
& 0 \leq \lambda_i \leq C/2 & i \in I \\
& 0 \leq \mu_j \leq M z_j & j \in J
\end{array}
$$

$$
\begin{array}{lll}
\min & \sum_{s,s'\in I} \lambda_s y_s \lambda_{s'} y_{s'} K(\mathbf{x}_s, \mathbf{x}_{s'}) + \sum_{t,t'\in J} \mu_t y_t \mu_{t'} y_{t'} K(\mathbf{x}_t, \mathbf{x}_{t'}) & \\
 & +2 \sum_{s\in I, t\in J} \lambda_s y_s \lambda_t y_t K(\mathbf{x}_s, \mathbf{x}_t) + C \sum_{i\in I} \xi_i & \\
\text{s.t.} & z_\ell \in \{0,1\} & \ell \in L \\
 & \mathbf{a}_\ell^\top z \geq b_\ell & \ell \in L \\
 & y_i \left(\sum_{s\in I} \lambda_s y_s K(\mathbf{x}_s, \mathbf{x}_i) + \sum_{t\in J} \mu_t y_t K(\mathbf{x}_t, \mathbf{x}_i) + \beta \right) \geq 1 - \xi_i & i \in I \\
 & y_j \left(\sum_{s\in I} \lambda_s y_s K(\mathbf{x}_s, \mathbf{x}_j) + \sum_{t\in J} \mu_t y_t K(\mathbf{x}_t, \mathbf{x}_j) + \beta \right) \geq 1 - M(1 - z_j) & j \in J \\
 & \xi_i \geq 0 & i \in I \\
 & 0 \leq \lambda_i \leq C/2 & i \in I \\
 & 0 \leq \mu_j \leq M z_j & j \in J
\end{array}
$$

(Partial) Dual: The kernel trick

$$
\begin{array}{lll}
\min & \sum_{s,s'\in I} \lambda_s y_s \lambda_{s'} y_{s'} K(\mathbf{x}_s, \mathbf{x}_{s'}) + \sum_{t,t'\in J} \mu_t y_t \mu_{t'} y_{t'} K(\mathbf{x}_t, \mathbf{x}_{t'}) & \\
 & +2 \sum_{s\in I, t\in J} \lambda_s y_s \lambda_t y_t K(\mathbf{x}_s, \mathbf{x}_t) + C \sum_{i\in I} \xi_i & \\
\text{s.t.} & z_\ell \in \{0,1\} & \ell \in L \\
 & \mathbf{a}_\ell^\top z \geq b_\ell & \ell \in L \\
 & y_i \left(\sum_{s\in I} \lambda_s y_s K(\mathbf{x}_s, \mathbf{x}_i) + \sum_{t\in J} \mu_t y_t K(\mathbf{x}_t, \mathbf{x}_i) + \beta \right) \geq 1 - \xi_i & i \in I \\
 & y_j \left(\sum_{s\in I} \lambda_s y_s K(\mathbf{x}_s, \mathbf{x}_j) + \sum_{t\in J} \mu_t y_t K(\mathbf{x}_t, \mathbf{x}_j) + \beta \right) \geq 1 - M(1 - z_j) & j \in J \\
 & \xi_i \geq 0 & i \in I \\
 & 0 \leq \lambda_i \leq C/2 & i \in I \\
 & 0 \leq \mu_j \leq M z_j & j \in J
\end{array}
$$

Parameters involved:

- C, to be tuned
- M, to be fixed

(Partial) Dual: The kernel trick

$$
\begin{array}{lll}
\min & \sum_{s,s' \in I} \lambda_s y_s \lambda_{s'} y_{s'} K(\mathbf{x}_s, \mathbf{x}_{s'}) + \sum_{t,t' \in J} \mu_t y_t \mu_{t'} y_{t'} K(\mathbf{x}_t, \mathbf{x}_{t'}) & \\
 & +2 \sum_{s \in I, t \in J} \lambda_s y_s \lambda_t y_t K(\mathbf{x}_s, \mathbf{x}_t) + C \sum_{i \in I} \xi_i & \\
\text{s.t.} & z_\ell \in \{0, 1\} & \ell \in L \\
 & \mathbf{a}_\ell^\top z \geq b_\ell & \ell \in L \\
 & y_i \left(\sum_{s \in I} \lambda_s y_s K(\mathbf{x}_s, \mathbf{x}_i) + \sum_{t \in J} \mu_t y_t K(\mathbf{x}_t, \mathbf{x}_i) + \beta \right) \geq 1 - \xi_i & i \in I \\
 & y_j \left(\sum_{s \in I} \lambda_s y_s K(\mathbf{x}_s, \mathbf{x}_j) + \sum_{t \in J} \mu_t y_t K(\mathbf{x}_t, \mathbf{x}_j) + \beta \right) \geq 1 - M(1 - z_j) & j \in J \\
 & \xi_i \geq 0 & i \in I \\
 & 0 \leq \lambda_i \leq C/2 & i \in I \\
 & 0 \leq \mu_j \leq M z_j & j \in J
\end{array}
$$

Parameters involved:

- C, to be tuned
- M, to be fixed?

(Partial) Dual: The kernel trick

$$
\begin{array}{lll}
\min & \sum_{s,s'\in I} \lambda_s y_s \lambda_{s'} y_{s'} K(\mathbf{x}_s, \mathbf{x}_{s'}) + \sum_{t,t'\in J} \mu_t y_t \mu_{t'} y_{t'} K(\mathbf{x}_t, \mathbf{x}_{t'}) & \\
& +2 \sum_{s\in I, t\in J} \lambda_s y_s \lambda_t y_t K(\mathbf{x}_s, \mathbf{x}_t) + C \sum_{i\in I} \xi_i & \\
\text{s.t.} & z_\ell \in \{0,1\} & \ell \in L \\
& \mathbf{a}_\ell^\top z \geq b_\ell & \ell \in L \\
& y_i \left(\sum_{s\in I} \lambda_s y_s K(\mathbf{x}_s, \mathbf{x}_i) + \sum_{t\in J} \mu_t y_t K(\mathbf{x}_t, \mathbf{x}_i) + \beta\right) \geq 1 - \xi_i & i \in I \\
& y_j \left(\sum_{s\in I} \lambda_s y_s K(\mathbf{x}_s, \mathbf{x}_j) + \sum_{t\in J} \mu_t y_t K(\mathbf{x}_t, \mathbf{x}_j) + \beta\right) \geq 1 - M(1-z_j) & j \in J \\
& \xi_i \geq 0 & i \in I \\
& 0 \leq \lambda_i \leq C/2 & i \in I \\
& 0 \leq \mu_j \leq M z_j & j \in J
\end{array}
$$

Parameters involved:

- C, to be tuned
- M, to be fixed?

Straightforward extension to several anchors

Experiments

- RBF kernel, parameters tuned by grid search
- Python + Gurobi
- $M = 100$, time.limit = 300 sec

Experiments

- RBF kernel, parameters tuned by grid search
- Python + Gurobi
- $M = 100$, time.limit = 300 sec

Data sets

Name	$\|\Omega\|$	V	$\|\Omega_+\|$	(%)
`wisconsin`	567	30	357	(62.7%)
`australian`	690	14	383	(55.5%)
`votes`	435	16	267	(61.4%)
`german`	1000	45	700	(70%)

Results. Increasing TNR (0.025)

Name		SVM		CSVM	
		Mean	Std	Mean	Std
wisconsin	TPR	0.990	0.017	0.945	0.045
	TNR	**0.948**	**0.049**	**0.965**	**0.037**
australian	TPR	0.863	0.079	0.772	0.081
	TNR	**0.830**	**0.071**	**0.903**	**0.050**
votes	TPR	0.963	0.040	0.846	0.097
	TNR	**0.951**	**0.031**	**0.978**	**0.038**
german	TPR	0.905	0.036	0.791	0.063
	TNR	**0.405**	**0.114**	**0.547**	**0.141**

Results. Increasing TPR (0.025)

Name		SVM		CSVM	
		Mean	Std	Mean	Std
`wisconsin`	`TPR`	**0.990**	**0.017**	**0.989**	**0.018**
	`TNR`	0.948	0.049	0.856	0.153
`australian`	`TPR`	**0.863**	**0.079**	**0.910**	**0.047**
	`TNR`	0.830	0.071	0.694	0.092
`votes`	`TPR`	**0.963**	**0.040**	**0.978**	**0.026**
	`TNR`	0.951	0.031	0.922	0.040
`german`					

Data sets

Name	$\|\Omega\|$	V	$\|\Omega_+\|$	(%)
`wisconsin`	567	30	357	(62.7%)
`australian`	690	14	383	(55.5%)
`votes`	435	16	267	(61.4%)
`german`	1000	45	700	(70%)

G-mean criterion

	SVM		CSVM (TNR$\geq$ 0.65)		CSVM (TNR $\geq$ 0.7)	
	Mean	Std	Mean	Std	Mean	Std
TPR	0.905	0.036	0.668	0.111	0.683	0.073
TNR	0.405	0.114	0.671	0.164	0.690	0.103

Feature selection

Benítez-Peña, Blanquero, C., Ramírez-Cobo, Cost-sensitive feature selection for support vector machines. Computers & OR, 2019.

Aim

- Find a minimum-cost (e.g. minimum-cardinality) set of features
  - Attaining $\widehat{\pi}_\ell(\omega, \beta) \geq \gamma_\ell^*$, $\ell \in L$
  - Hoping $\pi_\ell(\omega, \beta; I) \geq \gamma_\ell$, $\ell \in L$
- Once identified the features, solve an SVM

Feature selection. Linear kernel

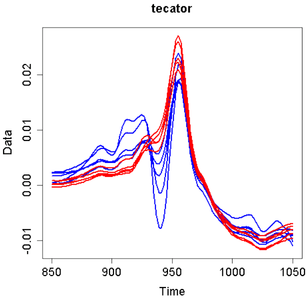
$$
\begin{array}{lll}
\min_{\boldsymbol{w},\beta,z,\zeta} & \displaystyle\sum_{k=1}^{N} \delta_k z_k & \\
s.t. & y_i(\boldsymbol{w}^\top x_i + \beta) \geq 1 - L(1-\zeta_i), & \forall i \in I \\
& \sum_{i\in I} \zeta_i(1-y_i) \geq \lambda_{-1} \sum_{i\in I}(1-y_i) & \\
& \sum_{i\in I} \zeta_i(1+y_i) \geq \lambda_{1} \sum_{i\in I}(1+y_i) & \\
& |w_k| \leq M z_k & \forall k \in 1,\ldots,N \\
& \zeta_i \in \{0,1\} & \forall i \in I \\
& z_k \in \{0,1\} & \forall k \in 1,\ldots,N
\end{array}
$$

Results. Linear kernel

Name		SVM		FS		Reduction
		Mean	Std	Mean	Std	
`wisconsin`	`TPR`	`0.992`	`0.013`	`0.975`	`0.023`	$30 \to 6.2$ (0.919 Std)
	`TNR`	`0.943`	`0.051`	`0.947`	`0.048`	
`votes`	`TPR`	`0.955`	`0.038`	`0.96`	`0.034`	$32 \to 9.3$ (1.16 Std)
	`TNR`	`0.947`	`0.059`	`0.945`	`0.052`	
`nursery`	`TPR`	`1`	`0`	`1`	`0`	$19 \to 1$ (0 Std)
	`TNR`	`1`	`0`	`1`	`0`	
`australian`	`TPR`	`0.769`	`0.083`	`0.772`	`0.074`	$34 \to 5.75$ (1.89 Std)
	`TNR`	`0.912`	`0.05`	`0.924`	`0.053`	
`careval`	`TPR`	`0.96`	`0.022`	`0.962`	`0.018`	$15 \to 11$ (0 Std)
	`TNR`	`0.948`	`0.024`	`0.935`	`0.039`	

Results. Radial kernel

Classification with functional data



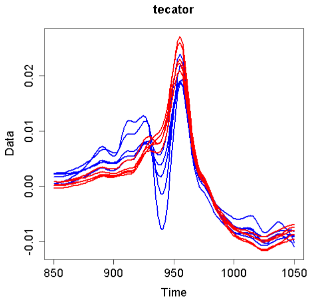
- $\mathbf{x} \in \mathcal{C}^0([0,T])$

Ferraty and Vieu. *Nonparametric functional data analysis: theory and practice*, 2006.

Ramsay and Silverman. *Functional data analysis*, 2006.

Febrero-Bande and Oviedo de la Fuente. "Statistical computing in functional data analysis: the r package fda.usc". *Journal of Statistical Software*, 2012.

Classification with functional data



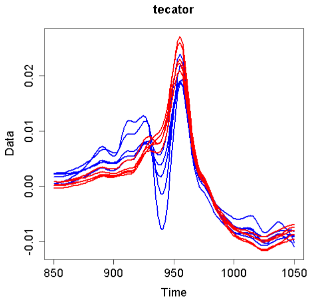
- $\mathbf{x} \in \mathcal{C}^0([0,T])$
- $\mathbf{x} \approx (\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m)) \in \mathbb{R}^m$

Ferraty and Vieu. *Nonparametric functional data analysis: theory and practice*, 2006.

Ramsay and Silverman. *Functional data analysis*, 2006.

Febrero-Bande and Oviedo de la Fuente. "Statistical computing in functional data analysis: the r package fda.usc". *Journal of Statistical Software*, 2012.

Classification with functional data



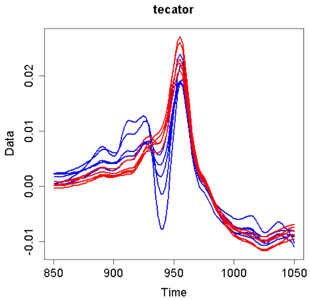
- $\mathbf{x} \in \mathcal{C}^0([0,T])$
- $\mathbf{x} \approx (\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m)) \in \mathbb{R}^m$
- $(\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m)) \approx \mathbf{x} \in \mathcal{C}^0([0,T])$

Ferraty and Vieu. *Nonparametric functional data analysis: theory and practice*, 2006.

Ramsay and Silverman. *Functional data analysis*, 2006.

Febrero-Bande and Oviedo de la Fuente. "Statistical computing in functional data analysis: the r package fda.usc". *Journal of Statistical Software*, 2012.

Classification with functional data



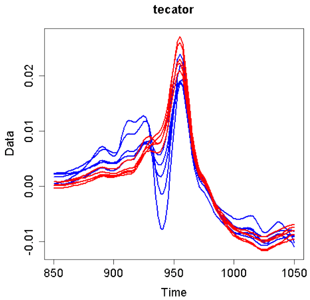
- $\mathbf{x} \in \mathcal{C}^0([0,T])$
- $\mathbf{x} \approx (\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m)) \in \mathbb{R}^m$
- $(\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m)) \approx \mathbf{x} \in \mathcal{C}^0([0,T]) \approx (\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m))$

Ferraty and Vieu. *Nonparametric functional data analysis: theory and practice*, 2006.

Ramsay and Silverman. *Functional data analysis*, 2006.

Febrero-Bande and Oviedo de la Fuente. "Statistical computing in functional data analysis: the r package fda.usc". *Journal of Statistical Software*, 2012.

Classification with functional data

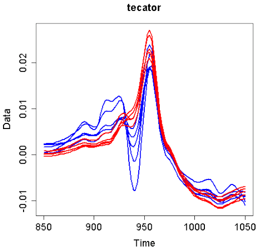


- $\mathbf{x} \in \mathcal{C}^0([0,T])$
- $\mathbf{x} \approx (\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m)) \in \mathbb{R}^m$
- $(\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m)) \approx \mathbf{x} \in \mathcal{C}^0([0,T]) \approx (\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m))$

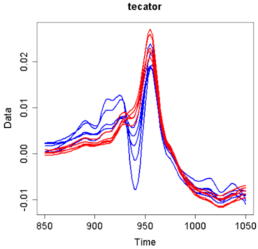
Muñoz and González. "Representing functional data using support vector machines". *Pattern Recognition Letters*, 2010.

Rossi and Villa. "Support vector machine for functional data classification". *Neurocomputing*, 2006.

Gaussian kernel for functional data (I)



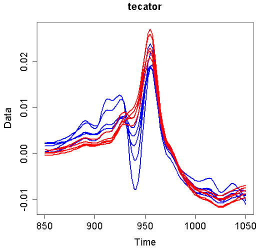
Gaussian kernel for functional data (I)



$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}$$

- $\|\mathbf{x}_i - \mathbf{x}_j\|^2 = \int_0^T (\mathbf{x}_i(t) - \mathbf{x}_j(t))^2 \, dt$

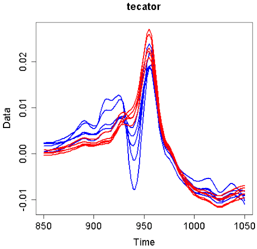
Gaussian kernel for functional data (I)



$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}$$

- $\|\mathbf{x}_i - \mathbf{x}_j\|^2 = \int_0^T (\mathbf{x}_i(t) - \mathbf{x}_j(t))^2 \, dt \approx \sum_k \eta_k (\mathbf{x}_i(t_k) - \mathbf{x}_j(t_k))^2$

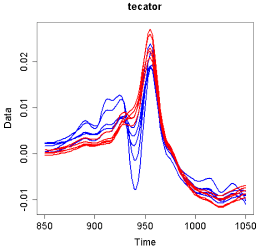
Gaussian kernel for functional data (I)



$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}$$

- $\|\mathbf{x}_i - \mathbf{x}_j\|^2 = \int_0^T (\mathbf{x}_i(t) - \mathbf{x}_j(t))^2 \, dt \approx \sum_k \eta_k (\mathbf{x}_i(t_k) - \mathbf{x}_j(t_k))^2$
- $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\int_0^T \gamma(\mathbf{x}_i(t) - \mathbf{x}_j(t))^2 \, dt} \approx e^{-\sum_k \gamma \eta_k (\mathbf{x}_i(t_k) - \mathbf{x}_j(t_k))^2}$

Gaussian kernel for functional data (I)



$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}$$

- $\|\mathbf{x}_i - \mathbf{x}_j\|^2 = \int_0^T (\mathbf{x}_i(t) - \mathbf{x}_j(t))^2 \, dt \approx \sum_k \eta_k (\mathbf{x}_i(t_k) - \mathbf{x}_j(t_k))^2$
- $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\int_0^T \gamma (\mathbf{x}_i(t) - \mathbf{x}_j(t))^2 \, dt} \approx e^{-\sum_k \gamma \eta_k (\mathbf{x}_i(t_k) - \mathbf{x}_j(t_k))^2}$

Gaussian kernel with functional bandwidth

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\int_0^T \gamma(t) (\mathbf{x}_i(t) - \mathbf{x}_j(t))^2 \, dt}$$

Gaussian kernel with functional bandwidth

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\int_0^T \gamma(t)(\mathbf{x}_i(t)-\mathbf{x}_j(t))^2\, dt}$$

A possible model for γ

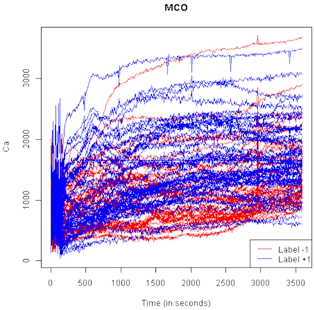
$$\gamma(t) = \begin{cases} \gamma_1, & \text{if } 0 \leq t \leq \tau_1 \\ \gamma_2, & \text{if } \tau_1 < t \leq \tau_2 \\ \ldots & \ldots \\ \gamma_H, & \text{if } \tau_{H-1} < t \leq T \end{cases}$$

- $\gamma_1, \ldots, \gamma_H \geq 0$
- $0 \leq \tau_1 \leq \ldots \leq \tau_{H-1} \leq T$

Blanquero, C., Jiménez-Cordero, Martín-Barragán. Functional-bandwidth kernel for Support Vector Machine with Functional Data: An alternating optimization algorithm. EJOR, 2019

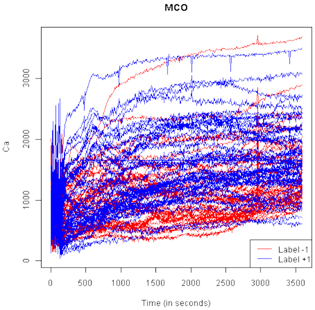
An example: Mitochondrial calcium data set

- 360 time instants in $[0, T]$, $T = 3590$
- 44 mice in treatment (+1), 45 control (-1)



An example: Mitochondrial calcium data set

- 360 time instants in $[0, T]$, $T = 3590$
- 44 mice in treatment (+1), 45 control (-1)



Out of sample accuracy estimates

$$\gamma(t) = \gamma \qquad \gamma(t) = \begin{cases} \gamma_1, & \text{if } 0 \leq t \leq \tau_1 \\ \gamma_2, & \text{if } \tau_1 < t \leq T \end{cases}$$

	-1	$+1$	-1	$+1$
-1 :	37.55%	10.96%	42.58%	7.56%
$+1$	12.09%	35.61%	7.23%	42.95%

Parameters tuning (basic gaussian kernel)

$$
\begin{array}{lll}
\max_{\lambda} & \sum_{i \in I} \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i y_i \lambda_j y_j e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2} & \\
\text{s.t.} & \sum_{i \in I} \lambda_i y_i = 0 & \\
& 0 \leq \lambda_i \leq \frac{C}{2} & i \in I
\end{array}
\qquad (P_{I,C,\gamma})
$$

C, γ: k-fold crossvalidation

- I : split in k blocks of similar size, $I_1, \ldots, I_k$
- for each pair C, γ in a grid (e.g. $2^{-12} \ldots, 2^{12}$), estimate $acc(C, \gamma)$:
    - for each $i = 1, \ldots, k$
        - solve $(P_{I \setminus I_i, C, \gamma})$, yielding λ^i, β (via KKT)
        - calculate $acc_i(C, \gamma)$, fraction of correctly classified in I_i if classifier with λ^i, β were used
    - $acc(C, \gamma) = \frac{1}{k} \sum_{i=1}^{k} acc_i(C, \gamma)$

Parameters tuning (basic gaussian kernel)

$$
\begin{array}{lll}
\max_\lambda & \sum_{i\in I}\lambda_i - \frac{1}{2}\sum_{ij}\lambda_i y_i \lambda_j y_j e^{-\gamma\|\mathbf{x}_i-\mathbf{x}_j\|^2} & \\
\text{s.t.} & \sum_{i\in I}\lambda_i y_i = 0 & \\
 & 0 \leq \lambda_i \leq \frac{C}{2} & i \in I
\end{array}
\qquad (P_{I,C,\gamma})
$$

C, γ: k-fold crossvalidation

- I : split in k blocks of similar size, $I_1, \dots, I_k$
- for each pair C, γ in a grid (e.g. $2^{-12} \dots, 2^{12}$), estimate $acc(C, \gamma)$:
  - for each $i = 1, \dots, k$
    - solve $(P_{I\setminus I_i, C, \gamma})$, yielding λ^i, β (via KKT)
    - calculate $acc_i(C, \gamma)$, fraction of correctly classified in I_i if classifier with λ^i, β were used
  - $acc(C, \gamma) = \frac{1}{k}\sum_{i=1}^{k} acc_i(C, \gamma)$

Unfeasible for functional bandwidth kernel!!!

Parameters tuning (functional bandwidth kernel)

$\theta = (\gamma_1, \dots, \gamma_H | \tau_1, \dots, \tau_{H-1})$

Parameters tuning (functional bandwidth kernel)

$$\theta = (\gamma_1, \ldots, \gamma_H | \tau_1, \ldots, \tau_{H-1})$$

$$\widehat{y}_{I,C,\theta}(\mathbf{x}) = \sum_{i \in I} y_i \lambda_i^C K_\theta(\mathbf{x}, \mathbf{x}_i) + \beta^C$$

Parameters tuning (functional bandwidth kernel)

$$\theta = (\gamma_1, \dots, \gamma_H | \tau_1, \dots, \tau_{H-1}) \qquad \widehat{y}_{I,C,\theta}(\mathbf{x}) = \sum_{i \in I} y_i \lambda_i^C K_\theta(\mathbf{x}, \mathbf{x}_i) + \beta^C$$

$$\begin{array}{lll} \max_\lambda & \sum_{i \in I} \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i y_i \lambda_j y_j K_\theta(\mathbf{x}_i, \mathbf{x}_j) & \\ \text{s.t.} & \sum_{i \in I} \lambda_i y_i = 0 & \\ & 0 \leq \lambda_i \leq \frac{C}{2} & i \in I \end{array} \qquad (P_{I,C,\theta})$$

Parameters tuning (functional bandwidth kernel)

$$\theta = (\gamma_1, \ldots, \gamma_H | \tau_1, \ldots, \tau_{H-1}) \qquad \qquad \widehat{y}_{I,C,\theta}(\mathbf{x}) = \sum_{i \in I} y_i \lambda_i^C K_\theta(\mathbf{x}, \mathbf{x}_i) + \beta^C$$

$$\begin{array}{lll} \max_\lambda & \sum_{i \in I} \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i y_i \lambda_j y_j K_\theta(\mathbf{x}_i, \mathbf{x}_j) & \\ \text{s.t.} & \sum_{i \in I} \lambda_i y_i = 0 & \\ & 0 \leq \lambda_i \leq \frac{C}{2} & i \in I \end{array} \qquad (P_{I,C,\theta})$$

Randomly split sample I into I_1, I_2 and I_3. **for** C *in grid* **do**

end

Alternating Procedure repeat

1. Fixed θ, find λ^C solving $(P_{I_1,C,\theta})$

until

;

2. Fixed λ, find θ maximizing correlation of y and $\widehat{y}_{I,C,\theta}$ in I_2.
stopping criteria

Return as C the one with best misclassification rate in I_3.

Parameters tuning (functional bandwidth kernel)

$\theta = (\gamma_1, \dots, \gamma_H | \tau_1, \dots, \tau_{H-1})$

$$\widehat{y}_{I,C,\theta}(\mathbf{x}) = \sum_{i \in I} y_i \lambda_i^C K_\theta(\mathbf{x}, \mathbf{x}_i) + \beta^C$$

$$\begin{array}{lll} \max_\lambda & \sum_{i \in I} \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i y_i \lambda_j y_j K_\theta(\mathbf{x}_i, \mathbf{x}_j) & \\ \text{s.t.} & \sum_{i \in I} \lambda_i y_i = 0 & \\ & 0 \leq \lambda_i \leq \frac{C}{2} & i \in I \end{array} \qquad (P_{I,C,\theta})$$

Randomly split sample I into I_1, I_2 and I_3. **for** C *in grid* **do**

end

Alternating Procedure repeat

1. Fixed θ, find λ^C solving $(P_{I_1,C,\theta})$

until

;

2. Fixed λ, find θ maximizing correlation of y and $\widehat{y}_{I,C,\theta}$ in I_2.
stopping criteria

Return as C the one with best misclassification rate in I_3.

$$\widehat{y}_{I,C,\theta}(\mathbf{x}) = \sum_{i \in I} y_i \lambda_i^{C} K_{\theta}(\mathbf{x}, \mathbf{x}_i) + \beta$$

$$K_{\theta}(\mathbf{x}, \mathbf{x}_i) = e^{-\sum_{h=1}^{H} \int_{\tau_{h-1}}^{\tau_h} \gamma_h (\mathbf{x}(t) - \mathbf{x}_i(t))^2 \, dt}$$

$$\widehat{y}_{I,C,\theta}(\mathbf{x}) = \sum_{i \in I} y_i \lambda_i^C K_\theta(\mathbf{x}, \mathbf{x}_i) + \beta$$

$$K_\theta(\mathbf{x}, \mathbf{x}_i) = e^{-\sum_{h=1}^{H} \int_{\tau_{h-1}}^{\tau_h} \gamma_h (\mathbf{x}(t) - \mathbf{x}_i(t))^2 \, dt}$$

Smooth optimization problem

- chain rule
- $K : \mathcal{C}^1$ for $\mathbf{x} : \mathcal{C}^0$
- $K : \mathcal{C}^3$ for $\mathbf{x} : \mathcal{C}^2$ (as generated by cubic spline)

Parameters tuning (functional bandwidth kernel)
Improved

Model H

$$\gamma(t) = \begin{cases} \gamma_1, & \text{if } 0 \leq t \leq \tau_1 \\ \gamma_2, & \text{if } \tau_1 < t \leq \tau_2 \\ \cdots & \cdots \\ \gamma_H, & \text{if } \tau_{H-1} < t \leq T \end{cases}$$

Nested heuristic

C., Martín-Barragán, Romero Morales, *Computers & OR*, 2014

Parameters tuning (functional bandwidth kernel)
Improved

Model H

$$\gamma(t) = \begin{cases} \gamma_1, & \text{if } 0 \leq t \leq \tau_1 \\ \gamma_2, & \text{if } \tau_1 < t \leq \tau_2 \\ \dots & \dots \\ \gamma_H, & \text{if } \tau_{H-1} < t \leq T \end{cases}$$

Nested heuristic

C., Martín-Barragán, Romero Morales, *Computers & OR*, 2014

Model 1

$\gamma(t) = \gamma_1$

Parameters tuning (functional bandwidth kernel)
Improved

Model H

$$\gamma(t) = \begin{cases} \gamma_1, & \text{if } 0 \le t \le \tau_1 \\ \gamma_2, & \text{if } \tau_1 < t \le \tau_2 \\ \cdots & \cdots \\ \gamma_H, & \text{if } \tau_{H-1} < t \le T \end{cases}$$

Nested heuristic

C., Martín-Barragán, Romero Morales, *Computers & OR*, 2014

Model 1

$\gamma(t) = \gamma_1$

Model 2

$$\gamma(t) = \begin{cases} \gamma_1, & \text{if } 0 \le t \le \tau_1 \\ \gamma_2, & \text{if } \tau_1 < t \le T \end{cases}$$

Parameters tuning (functional bandwidth kernel)
Improved

Model H

$$\gamma(t) = \begin{cases} \gamma_1, & \text{if } 0 \leq t \leq \tau_1 \\ \gamma_2, & \text{if } \tau_1 < t \leq \tau_2 \\ \cdots & \cdots \\ \gamma_H, & \text{if } \tau_{H-1} < t \leq T \end{cases}$$

Nested heuristic

C., Martín-Barragán, Romero Morales, *Computers & OR*, 2014

Model 1

$\gamma(t) = \gamma_1$

Model 2

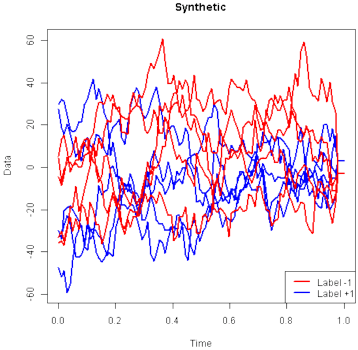
$$\gamma(t) = \begin{cases} \gamma_1, & \text{if } 0 \leq t \leq \tau_1 \\ \gamma_2, & \text{if } \tau_1 < t \leq T \end{cases}$$

Model 3

$$\gamma(t) = \begin{cases} \gamma_1, & \text{if } 0 \leq t \leq \tau_1 \\ \gamma_2, & \text{if } \tau_1 < t \leq \tau_2 \\ \gamma_3, & \text{if } \tau_2 < t \leq T \end{cases} \quad \cdots$$

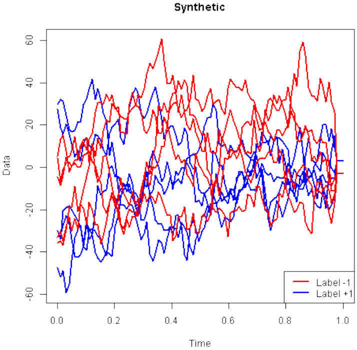
A test example

15,000 functions like these



A test example

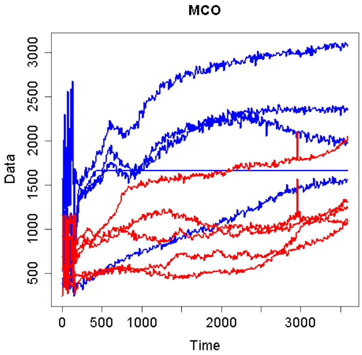
15,000 functions like these



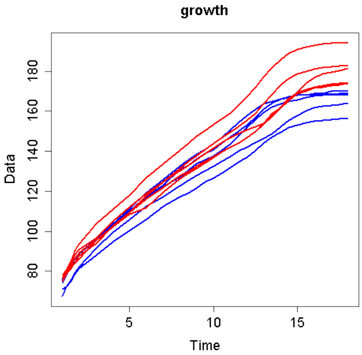
	1 (classic SVM)	$H = 2$	$H = 3$	$H = 4$
% misc	32.95	0	0	0

	#records	#time instants	#records label -1	#records label +1
`MCO`	89	360	44	45
`growth`	93	31	54	39
`phoneme`	200	150	100	100
`rain`	35	365	15	20
`regions`	35	365	20	15
`tecator`	215	100	77	138

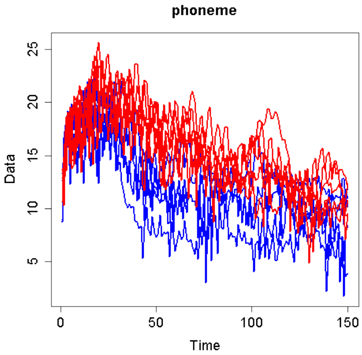
	#records	#time instants	#records label -1	#records label +1
MCO	89	360	44	45
growth	93	31	54	39
phoneme	200	150	100	100
rain	35	365	15	20
regions	35	365	20	15
tecator	215	100	77	138



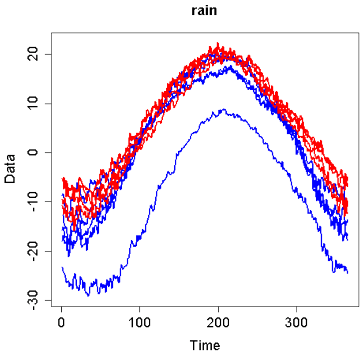
	#records	#time instants	#records label -1	#records label +1
MCO	89	360	44	45
growth	93	31	54	39
phoneme	200	150	100	100
rain	35	365	15	20
regions	35	365	20	15
tecator	215	100	77	138



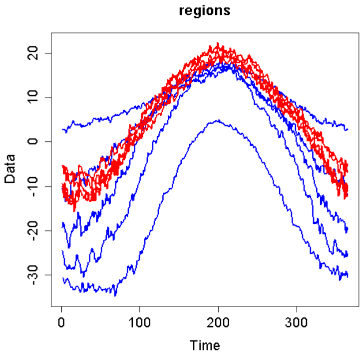
	#records	#time instants	#records label -1	#records label +1
`MCO`	89	360	44	45
`growth`	93	31	54	39
`phoneme`	200	150	100	100
`rain`	35	365	15	20
`regions`	35	365	20	15
`tecator`	215	100	77	138



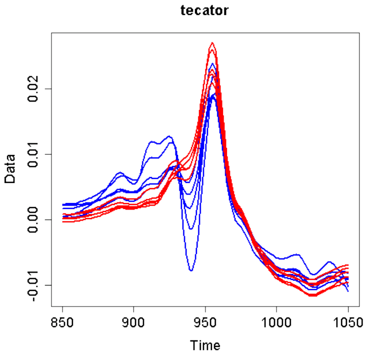
	#records	#time instants	#records label -1	#records label +1
`MCO`	89	360	44	45
`growth`	93	31	54	39
`phoneme`	200	150	100	100
`rain`	35	365	15	20
`regions`	35	365	20	15
`tecator`	215	100	77	138



	#records	#time instants	#records label -1	#records label +1
MCO	89	360	44	45
growth	93	31	54	39
phoneme	200	150	100	100
rain	35	365	15	20
regions	35	365	20	15
tecator	215	100	77	138



	#records	#time instants	#records label -1	#records label +1
MCO	89	360	44	45
growth	93	31	54	39
phoneme	200	150	100	100
rain	35	365	15	20
regions	35	365	20	15
tecator	215	100	77	138

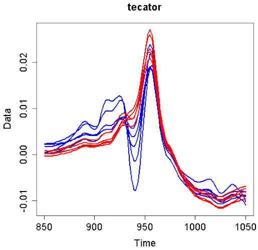


	#records	#time instants	#records label -1	#records label +1
`MCO`	89	360	44	45
`growth`	93	31	54	39
`phoneme`	200	150	100	100
`rain`	35	365	15	20
`regions`	35	365	20	15
`tecator`	215	100	77	138

% misclassification rate (out-of-sample)

	$H = 1$	$H = 2$	$H = 3$	$H = 4$
`MCO`	20.80	14.73	11.05	10.37
`growth`	5.64	4.67	4.35	4.19
`phoneme`	19.88	18.08	17.63	17.11
`rain`	28.40	22.84	22.42	21.59
`regions`	19.46	16.43	16.02	16.51
`tecator`	3.47	2.92	2.64	2.29

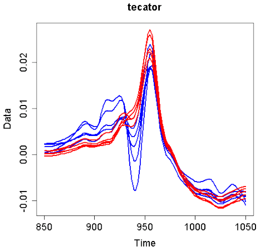
Gaussian kernel for functional data (II)



$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}$$

- $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\int_0^T \gamma (\mathbf{x}_i(t) - \mathbf{x}_j(t))^2 \, dt}$

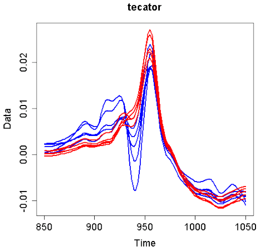
Gaussian kernel for functional data (II)



$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}$$

- $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\int_0^T \gamma (\mathbf{x}_i(t) - \mathbf{x}_j(t))^2 \, dt}$
- $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\int_0^T \gamma(t) (\mathbf{x}_i(t) - \mathbf{x}_j(t))^2 \, dt}$

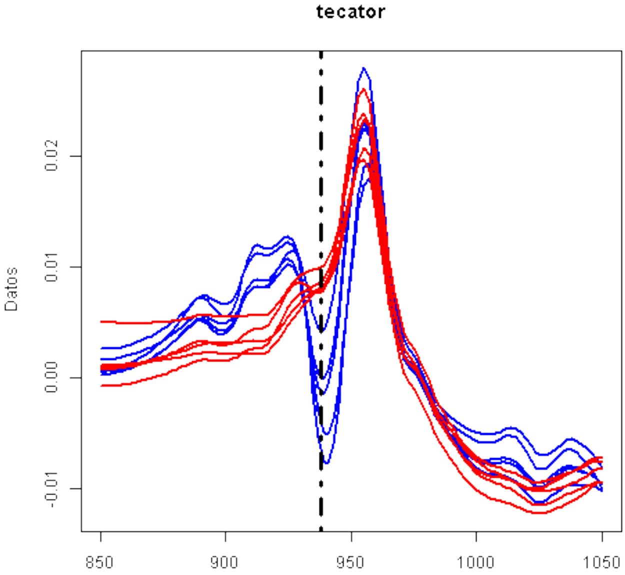
Gaussian kernel for functional data (II)



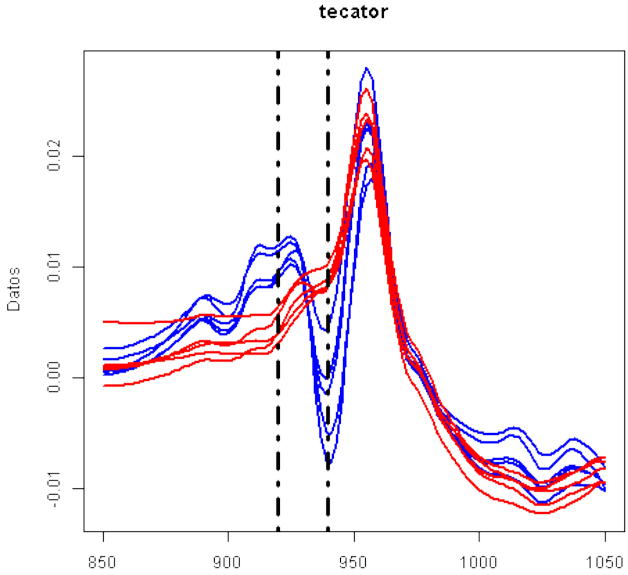
$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}$$

- $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\int_0^T \gamma (\mathbf{x}_i(t) - \mathbf{x}_j(t))^2 \, dt}$
- $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\int_0^T \gamma(t) (\mathbf{x}_i(t) - \mathbf{x}_j(t))^2 \, dt}$
- $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\sum_{h=1}^{H} \gamma (\mathbf{x}_i(\tau_h) - \mathbf{x}_j(\tau_h))^2}$
  - $\gamma \geq 0$
  - $0 \leq \tau_{h-1} \leq \tau_h - \delta$, $h = 1, \ldots, H$

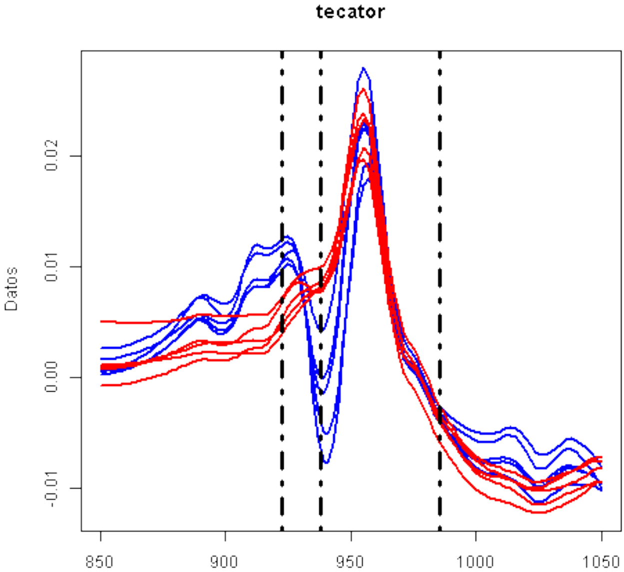
SVM with functional data



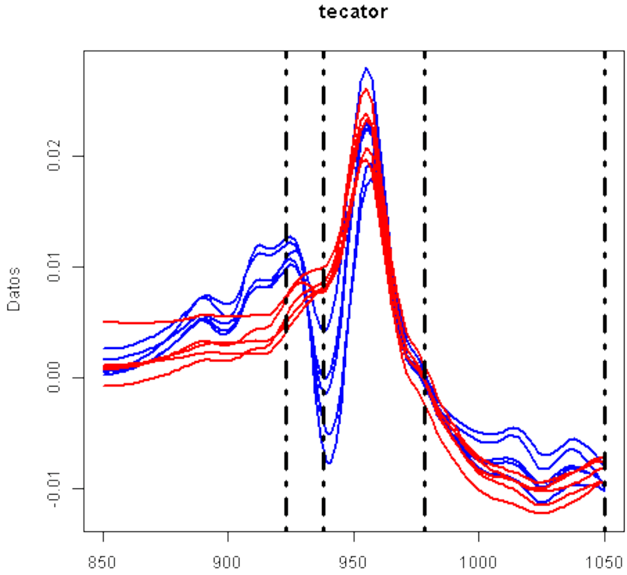
SVM with functional data



SVM with functional data



SVM with functional data



Parameters tuning (time instants selection)

Blanquero, C., Jiménez-Cordero, Martín-Barragán. Variable selection in classification for multivariate functional data. Information Sciences, 2019.

$$\theta = (\gamma | \tau_1, \ldots, \tau_{H-1})$$

Parameters tuning (time instants selection)

Blanquero, C., Jiménez-Cordero, Martín-Barragán. Variable selection in classification for multivariate functional data. Information Sciences, 2019.

$$\theta = (\gamma | \tau_1, \ldots, \tau_{H-1}) \qquad \widehat{y}_{I,C,\theta}(\mathbf{x}) = \sum_{i \in I} y_i \lambda_i^C K_\theta(\mathbf{x}, \mathbf{x}_i) + \beta^C$$

Randomly split sample I into I_1, I_2 and I_3. **for** C *in grid* **do**

end

Alternating Procedure repeat

1. Fixed θ, find λ^C solving $(P_{I_1,C,\theta})$

until

;

2. Fixed λ, find θ maximizing correlation of y and $\widehat{y}_{I,C,\theta}$ in I_2.
stopping criteria

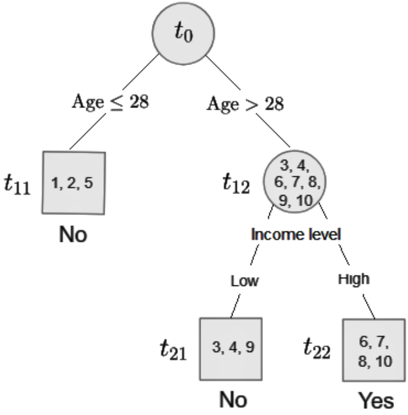
Return as C the one with best misclassification rate in I_3.
Return as λ and θ those associated with C

% misclassification rate (out-of-sample)

	SVM	$H = 1$	$H = 2$	$H = 3$	$H = 4$
MCO	20.80	29.02	18.64	18.14	18.81
$growth$	5.64	13.22	4.67	4.03	3.87
$phoneme$	19.88	18.00	16.96	16.36	16.20
$rain$	28.40	10.75	11.66	11.66	10
$regions$	19.46	20.75	10.26	8.10	7.23
$tecator$	3.47	4.66	2.22	2.08	1.52

CARTs (Breiman et al. 1984)

Applicant	Age	Income level	Loan granted
1	22	Low	No
2	26	High	No
3	30	Low	Yes
4	32	Low	No
5	20	High	No
6	45	High	Yes
7	60	High	No
8	54	High	Yes
9	50	Low	No
10	48	High	Yes



Motivation

Pros

- They are rule-based and, when they are not very deep, deemed to be easy-to-interpret.
- Low computational times.

Cons

- Classification Trees is a GREEDY procedure, not OPTIMAL.

$+$ Advances in both computer performance and Mathematical Optimization solvers

literature

- Integer Programming-based strategies:
  + Bertsimas and Dunn 2017.
  + Bertsimas, Dunn and Mundru, 2019.
  + Günlük et al. 2018.
  + Verwer and Zhang 2017, Verwer et al. 2017.
- It is commonly assumed that training sets are small.
- A CPU time limit is imposed to the solver.

literature

- Integer Programming-based strategies:
  - Bertsimas and Dunn 2017.
  - Bertsimas, Dunn and Mundru, 2019.
  - Günlük et al. 2018.
  - Verwer and Zhang 2017, Verwer et al. 2017.
- It is commonly assumed that training sets are small.
- A CPU time limit is imposed to the solver.

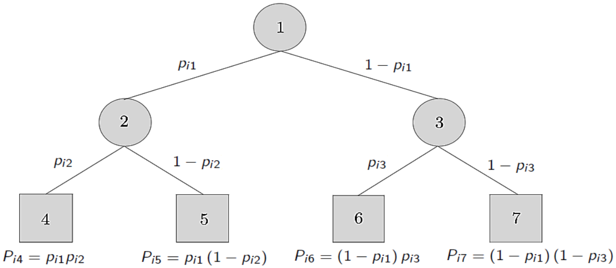
Our proposal: a **continuous** optimization-based method which yields **better results** by performing several local searches in relatively **short time**.

Optimal Randomized Classification Trees

We have a sample $I = \{(\boldsymbol{x}_i, y_i)\}_{1 \leq i \leq n}$, where $\boldsymbol{x}_i \in [0,1]^p$ and $y_i \in \{1, \ldots, K\}$.

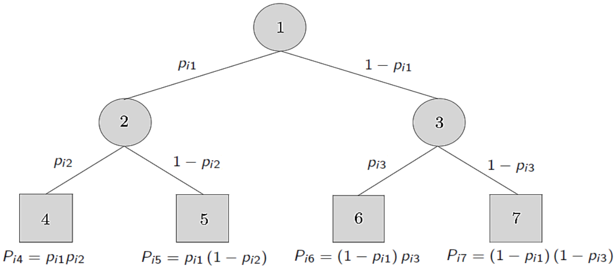
Optimal Randomized Classification Trees

We have a sample $I = \{(\boldsymbol{x}_i, y_i)\}_{1 \le i \le n}$, where $\boldsymbol{x}_i \in [0,1]^p$ and $y_i \in \{1, \ldots, K\}$.
A maximal binary tree of depth D. Nodes: Branch $t \in \tau_B$, Leaf $t \in \tau_L$.



Optimal Randomized Classification Trees

We have a sample $I = \{(\boldsymbol{x}_i, y_i)\}_{1 \leq i \leq n}$, where $\boldsymbol{x}_i \in [0,1]^p$ and $y_i \in \{1, \ldots, K\}$.
A maximal binary tree of depth D. Nodes: Branch $t \in \tau_B$, Leaf $t \in \tau_L$.



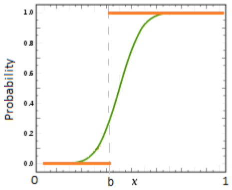
- Oblique splits:

$a_{jt} \in [-1, 1]$	coefficient of predictor variable j in the oblique cut over branch node $t \in \tau_B$,
$\mu_t \in [-1, 1]$	location parameter at branch node $t \in \tau_B$.

Optimal Randomized Classification Trees

- Probabilities

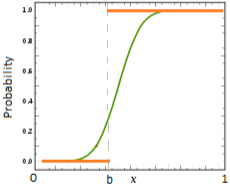
CDF $F(\cdot)$



Optimal Randomized Classification Trees

- Probabilities

CDF $F(\cdot)$

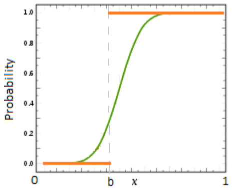


$$p_{it}\left(\boldsymbol{a}_{\cdot t}, \mu_t\right) = F\left(\frac{1}{p}\sum_{j=1}^{p} a_{jt}x_{ij} - \mu_t\right),\ i = 1, \ldots, n,\ t \in \tau_B.$$

Optimal Randomized Classification Trees

- Probabilities

CDF $F(\cdot)$



$$p_{it}\left(\boldsymbol{a}_{\cdot t}, \mu_t\right) = F\left(\frac{1}{p}\sum_{j=1}^{p} a_{jt}x_{ij} - \mu_t\right),\ i = 1, \ldots, n,\ t \in \tau_B.$$

$$P_{it}\left(\boldsymbol{a}, \boldsymbol{\mu}\right) = \prod_{t_l \in N_L(t)} p_{it_l}\left(\boldsymbol{a}_{\cdot t_l}, \mu_{t_l}\right) \prod_{t_r \in N_R(t)} \left(1 - p_{it_r}\left(\boldsymbol{a}_{\cdot t_r}, \mu_{t_r}\right)\right),\ i = 1, \ldots, n,\ t \in \tau_L.$$

Optimal Randomized Classification Trees

- Each $t \in \tau_L$ is labeled with one class:

$$C_{kt} = \begin{cases} 1, & \text{node } t \text{ is labeled with class } k \\ 0, & \text{otherwise} \end{cases} \quad , k = 1, \dots, K, \; t \in \tau_L$$

$$\sum_{k=1}^{K} C_{kt} = 1, \; t \in \tau_L.$$

Optimal Randomized Classification Trees

- Each $t \in \tau_L$ is labeled with one class:

$$C_{kt} = \begin{cases} 1, & \text{node } t \text{ is labeled with class } k \\ 0, & \text{otherwise} \end{cases} \quad , k = 1, \ldots, K, \; t \in \tau_L$$

$$\sum_{k=1}^{K} C_{kt} = 1, \; t \in \tau_L.$$

- Each class $k = 1, \ldots, K$ is identified by, at least, one terminal node:

$$\sum_{t \in \tau_L} C_{kt} \geq 1, \; k = 1, \ldots, K.$$

- We now introduce a misclassification cost for classifying an individual from class k in class k':

$$W_{kk'} \geq 0, \ k, k' = 1, \ldots, K, \ k \neq k'.$$

Optimal Randomized Classification Trees

- We now introduce a misclassification cost for classifying an individual from class k in class k':

$$W_{kk'} \geq 0,\ \ k, k' = 1, \ldots, K,\ \ k \neq k'.$$

- **Objective**

$$\min \qquad \sum_{k=1}^{K} \sum_{i \in I_k} \sum_{t \in \tau_L} P_{it}\left(\boldsymbol{a}, \boldsymbol{\mu}\right) \sum_{k' \neq k} C_{k't} W_{kk'}$$

Optimal Randomized Classification Trees

(Mixed-Integer Non-Linear Optimization Problem)

$$
\begin{aligned}
\min \quad & \sum_{k=1}^{K} \sum_{i \in I_k} \sum_{t \in \tau_L} P_{it}\left(\boldsymbol{a}, \boldsymbol{\mu}\right) \sum_{k' \neq k} C_{k't} W_{kk'} \\
\text{s.t.} \quad & \sum_{k=1}^{K} C_{kt} = 1,\ t \in \tau_L, \\
& \sum_{t \in \tau_L} C_{kt} \geq 1,\ k = 1, \ldots, K, \\
& a_{jt} \in [-1, 1],\ j = 1, \ldots, p,\ t \in \tau_B, \\
& \mu_t \in [-1, 1],\ t \in \tau_B, \\
& C_{kt} \in \{0, 1\},\ k = 1, \ldots, K,\ t \in \tau_L.
\end{aligned}
$$

Optimal Randomized Classification Trees

(Continuous Non-Linear Optimization Problem)

$$
\begin{aligned}
\min \quad & \sum_{k=1}^{K} \sum_{i \in I_k} \sum_{t \in \tau_L} P_{it}\left(\boldsymbol{a}, \boldsymbol{\mu}\right) \sum_{k' \neq k} C_{k't} W_{kk'} \\
\text{s.t.} \quad & \sum_{k=1}^{K} C_{kt} = 1,\ t \in \tau_L, \\
& \sum_{t \in \tau_L} C_{kt} \geq 1,\ k = 1, \ldots, K, \\
& a_{jt} \in [-1, 1],\ j = 1, \ldots, p,\ t \in \tau_B, \\
& \mu_t \in [-1, 1],\ t \in \tau_B, \\
& C_{kt} \in [0, 1],\ k = 1, \ldots, K,\ t \in \tau_L.
\end{aligned}
\tag{ORCT}
$$

Theorem

There exists an optimal solution to ORCT such that $C_{kt} \in \{0, 1\}$, $k = 1, \ldots, K, t \in \tau_L$.

ORCT's prediction

A new unlabeled observation $\boldsymbol{x}$

$$\Downarrow$$

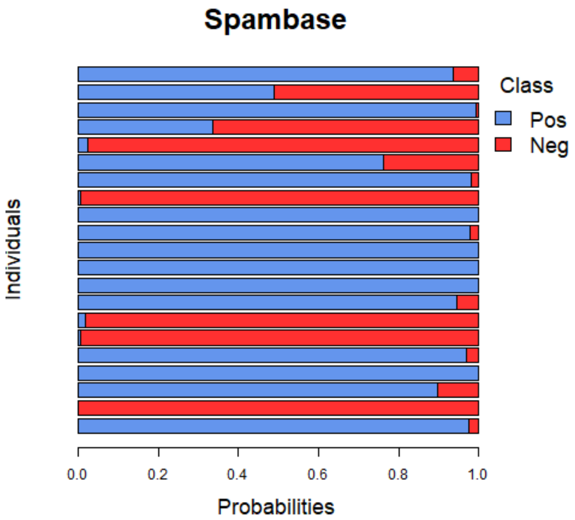
Once the optimization problem has been solved

,

the decision variables are used for predicting its class:

$$m_n(\boldsymbol{x}) = \arg\max_k \left\{ \sum_{t \in \tau_L} \mathbb{P}\left(\boldsymbol{x} \in k | \boldsymbol{x} \in t\right) \mathbb{P}\left(\boldsymbol{x} \in t\right) \right\} = \arg\max_k \left\{ \sum_{t \in \tau_L} C_{kt} \cdot P_{\boldsymbol{x}t}\left(\boldsymbol{a}, \boldsymbol{\mu}\right) \right\}.$$

ORCT's prediction



UCI Machine Learning Repository

Data set	n	p	K	Class distribution
Sonar	208	60	2	55% - 45%
Wisconsin	569	30	2	63% - 37%
Credit-approval	653	37	2	55% - 45%
Pima	768	8	2	65% - 35%
German-credit	1000	48	2	70% - 30%
Ozone	1848	72	2	97% - 3%
Spambase	4601	57	2	61% - 39%
Iris	150	4	3	33.3%-33.3%-33.3%
Wine	178	13	3	40%-33%-27%
Seeds	210	7	3	33.3%-33.3%-33.3%
Thyroid	3772	21	3	92.5%-5%-2.5%
Car	1728	15	4	70%-22%-4%-4%

Computational experience

- Logistic CDF:
$$F\left(\cdot;\gamma\right) = \frac{1}{1+\exp\left(-\left(\cdot\right)\gamma\right)},\ \ \gamma > 0,$$
$\gamma = 512.$

Computational experience

- Logistic CDF:
$$F\left(\cdot;\gamma\right)=\frac{1}{1+\exp\left(-\left(\cdot\right)\gamma\right)},\ \gamma>0,$$
$\gamma=512.$
- Equal misclassification weights,
$$W_{kk'}=0.5,\ k,k'=1,\ldots,K,\ k\neq k'.$$

Computational experience

- Logistic CDF:
$$F(\cdot;\gamma) = \frac{1}{1+\exp\left(-(\cdot)\,\gamma\right)},\ \ \gamma > 0,$$
$\gamma = 512.$
- Equal misclassification weights,
$$W_{kk'} = 0.5,\ \ k, k' = 1, \ldots, K,\ \ k \neq k'.$$
- 10 hold-out runs: training subset (75%) and test subset (25%).

Computational experience

- Logistic CDF:
$$F(\cdot;\gamma) = \frac{1}{1+\exp\left(-\left(\cdot\right)\gamma\right)}, \ \gamma > 0,$$
$\gamma = 512$.
- Equal misclassification weights,
$$W_{kk'} = 0.5, \ k, k' = 1, \dots, K, \ k \neq k'.$$
- 10 hold-out runs: training subset (75%) and test subset (25%).
- Performance measure: average accuracy over the 10 test subsets.

Computational experience

- Logistic CDF:

$$F(\cdot;\gamma) = \frac{1}{1+\exp\left(-\left(\cdot\right)\gamma\right)}, \ \gamma > 0,$$

  $\gamma = 512$.

- Equal misclassification weights,

$$W_{kk'} = 0.5, \ k, k' = 1, \dots, K, \ k \neq k'.$$

- 10 hold-out runs: training subset (75%) and test subset (25%).
- Performance measure: average accuracy over the 10 test subsets.
- Python 3.5, IPOPT 3.11.1 solver.

Computational experience

ORCT compared with:

- **CART** (Breiman et al. 1984).
- **OCT-H** (Bertsimas and Dunn 2017).

Computational experience

$D = 1$

Data set	ORCT average time (in secs)	Out-of-sample accuracy		
		ORCT	CART	OCT-H
Sonar	22	**76.3**	70.0	70.4
Wisconsin	24	**96.4**	92.0	93.1
Credit-approval	22	83.7	85.7	**87.9**
Pima	21	**75.8**	74.2	71.6
German-credit	28	**72.8**	72.1	71.6
Ozone	94	96.7	95.6	**96.8**
Spambase	72	**89.8**	89.2	83.6

Computational experience

$D = 1$

Data set	ORCT average time (in secs)	Out-of-sample accuracy		
		ORCT	CART	OCT-H
Sonar	22	**76.3**	70.0	70.4
Wisconsin	24	**96.4**	92.0	93.1
Credit-approval	22	83.7	85.7	**87.9**
Pima	21	**75.8**	74.2	71.6
German-credit	28	**72.8**	72.1	71.6
Ozone	94	96.7	95.6	**96.8**
Spambase	72	**89.8**	89.2	83.6

$D = 2$

Data set	ORCT average time (in secs)	Out-of-sample accuracy		
		ORCT	CART	OCT-H
Iris	17	**95.9**	92.7	95.1
Wine	23	**96.6**	88.6	91.1
Seeds	20	**94.2**	90.2	90.6
Thyroid	145	92.2	**99.1**	92.5
Car	71	**90.8**	88.1	87.5

Sparsity on ORCTs

$$
\begin{aligned}
\min \quad & \sum_{k=1}^{K} \sum_{i \in I_k} \sum_{t \in \tau_L} P_{it}\left(\boldsymbol{a}, \boldsymbol{\mu}\right) \sum_{k' \neq k} W_{kk'} C_{k't} \\
\text{s.t.} \quad & \sum_{k=1}^{K} C_{kt} = 1,\ t \in \tau_L, \\
& \sum_{t \in \tau_L} C_{kt} \geq 1,\ k = 1, \ldots, K, \\
& a_{jt} \in \left[-1, 1\right],\ j = 1, \ldots, p,\ t \in \tau_B, \\
& \mu_t \in \left[-1, 1\right],\ t \in \tau_B, \\
& C_{kt} \in \left[0, 1\right],\ k = 1, \ldots, K,\ t \in \tau_L,
\end{aligned}
$$

Sparsity on ORCTs

Local: less predictor variables at each node

$$
\begin{aligned}
\min \quad & \sum_{k=1}^{K} \sum_{i \in I_k} \sum_{t \in \tau_L} P_{it}\left(\boldsymbol{a}, \boldsymbol{\mu}\right) \sum_{k' \neq k} W_{kk'} C_{k't} \quad \boldsymbol{+ \lambda^L \sum_{j=1}^{p} \left\| a_{j\cdot} \right\|_1} \\
\text{s.t.} \quad & \sum_{k=1}^{K} C_{kt} = 1,\ t \in \tau_L, \\
& \sum_{t \in \tau_L} C_{kt} \geq 1,\ k = 1, \ldots, K, \\
& a_{jt} \in \left[-1, 1\right],\ j = 1, \ldots, p,\ t \in \tau_B, \\
& \mu_t \in \left[-1, 1\right],\ t \in \tau_B, \\
& C_{kt} \in \left[0, 1\right],\ k = 1, \ldots, K,\ t \in \tau_L,
\end{aligned}
$$

Sparsity on ORCTs

Local: less predictor variables at each node
Global: less predictor variables in the tree

$$\begin{aligned}
\min \quad & \sum_{k=1}^{K} \sum_{i \in I_k} \sum_{t \in \tau_L} P_{it}\left(\boldsymbol{a}, \boldsymbol{\mu}\right) \sum_{k' \neq k} W_{kk'} C_{k't} \; + \boldsymbol{\lambda^L} \sum_{\boldsymbol{j=1}}^{\boldsymbol{p}} \left\|\boldsymbol{a_{j\cdot}}\right\|_{\boldsymbol{1}} \; + \boldsymbol{\lambda^G} \sum_{\boldsymbol{j=1}}^{\boldsymbol{p}} \left\|\boldsymbol{a_{j\cdot}}\right\|_{\boldsymbol{\infty}} \\
\text{s.t.} \quad & \sum_{k=1}^{K} C_{kt} = 1,\ t \in \tau_L, \\
& \sum_{t \in \tau_L} C_{kt} \geq 1,\ k = 1, \ldots, K, \\
& a_{jt} \in \left[-1, 1\right],\ j = 1, \ldots, p,\ t \in \tau_B, \\
& \mu_t \in \left[-1, 1\right],\ t \in \tau_B, \\
& C_{kt} \in \left[0, 1\right],\ k = 1, \ldots, K,\ t \in \tau_L,
\end{aligned}$$

Theorem

Let $\sigma \in [0,1]$. **For**

$$\boldsymbol{\lambda^L \geq (1-\sigma) \max_{\substack{C_{kt}\in\{0,1\}\\ \mu_t\in[-1,1]}} \max_{j=1,\ldots,p} \left\|\nabla_{a_{j\cdot}} g\left(0,\mu,C\right)\right\|_\infty} \textbf{ and}$$

$$\boldsymbol{\lambda^G \geq \sigma \max_{\substack{C_{kt}\in\{0,1\}\\ \mu_t\in[-1,1]}} \max_{j=1,\ldots,p} \left\|\nabla_{a_{j\cdot}} g\left(0,\mu,C\right)\right\|_1},$$

$\boldsymbol{a=0}$ **is a stationary point** of the sparse ORCT, being g the misclassification cost term in the objective function.

Particular case

Sparse ORCT at depth 1 ($\lambda^L = \lambda^G$)

Theorem

Let $F \in \mathcal{C}^1$ a CDF with f as its corresponding PDF. **The minimum $\boldsymbol{\lambda}^L$ from which $\boldsymbol{a}_{\cdot 1} = \mathbf{0}$ is a stationary point to the sparse ORCT at depth 1 is**:

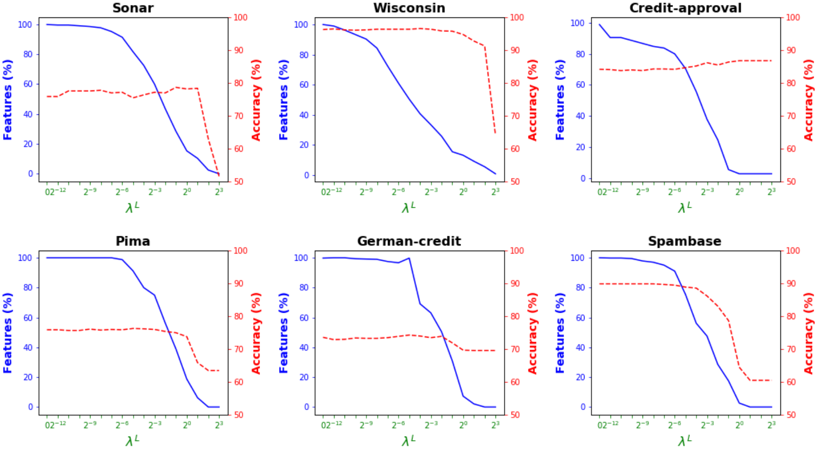
$$\boldsymbol{\lambda}^{\boldsymbol{L}} = \max\left\{\boldsymbol{\lambda}^{\boldsymbol{L}}_{\boldsymbol{\mu_1 = -1}}, \boldsymbol{\lambda}^{\boldsymbol{L}}_{\boldsymbol{\mu_1 = 1}}\right\},$$

where

$$\boldsymbol{\lambda}^{\boldsymbol{L}}_{\boldsymbol{\mu_1}} = \frac{\mathbf{1}}{\boldsymbol{p}} \boldsymbol{f}\left(-\frac{\boldsymbol{\mu_1}}{\boldsymbol{p}}\right) \max_{\boldsymbol{j=1,\ldots,p}} \left| -\boldsymbol{W_{21}} \sum_{\boldsymbol{i \in I_2}} \boldsymbol{x_{ij}} + \boldsymbol{W_{12}} \sum_{\boldsymbol{i \in I_1}} \boldsymbol{x_{ij}} \right|.$$

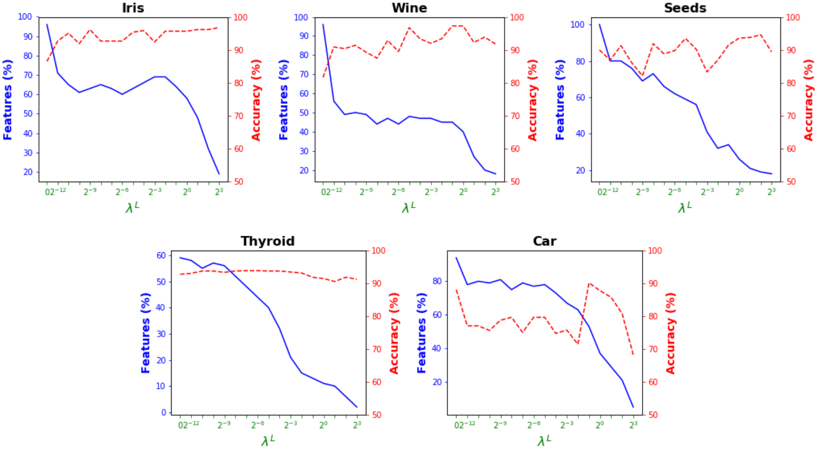
Results for local sparsity ($D = 1$)

λ^L varying, $\lambda^G = 0$



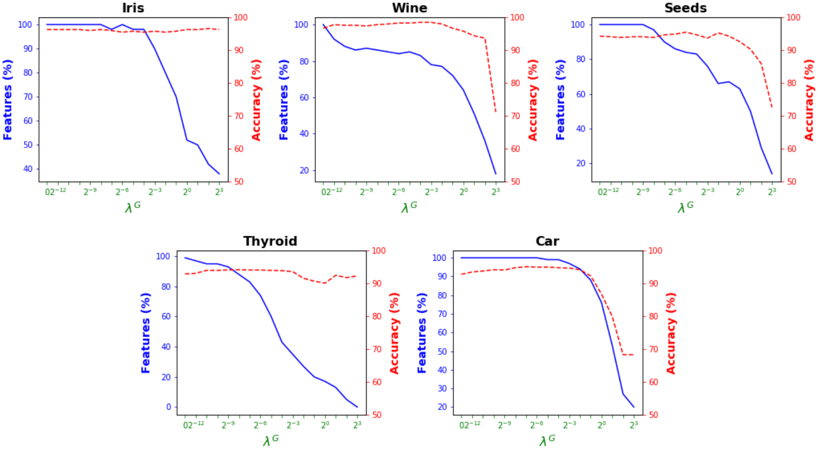
Results for local sparsity ($D = 2$)

λ^L varying, $\lambda^G = 0$



Results for global sparsity ($D = 2$)

$\lambda^L = 0$, λ^G varying



(Sparse) linear regression models

(Sparse and cost-sensitive) linear models using MINLO

$$\min_{\beta \in \mathcal{B}} \sum_{i=1}^{m} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2$$

$\mathcal{B}$ modelling, among other things, which features are selected

(Sparse and cost-sensitive) linear models using MINLO

$$\min_{\beta \in \mathcal{B}} \sum_{i=1}^{m} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2$$

$\mathcal{B}$ modelling, among other things, which features are selected

- Bertsimas and King, *Operations Research*, 2015.
- Bertsimas, King and Mazumder, *Annals of Statistics*, 2016.
- Bertsimas, Pauphilet, Van Parys, in `arXiv.org`, 2019.
- C., Olivares-Nadal, Ramírez-Cobo, *Biostatistics*, 2017.

$$Y_i = \sum_{j=1}^{N} \beta_j X_{ij} + e_i \qquad i = 1, \ldots, m$$

Sparsity in linear models via convex optim

$$Y_i = \sum_{j=1}^{N} \beta_j X_{ij} + e_i \qquad i = 1, \ldots, m \qquad\qquad \min_{\beta} \sum_{i=1}^{m} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2$$

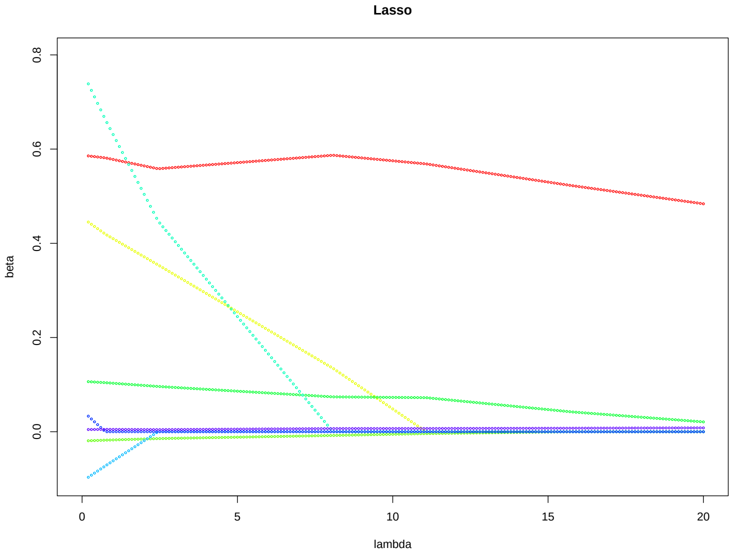
Sparsity in linear models via convex optim

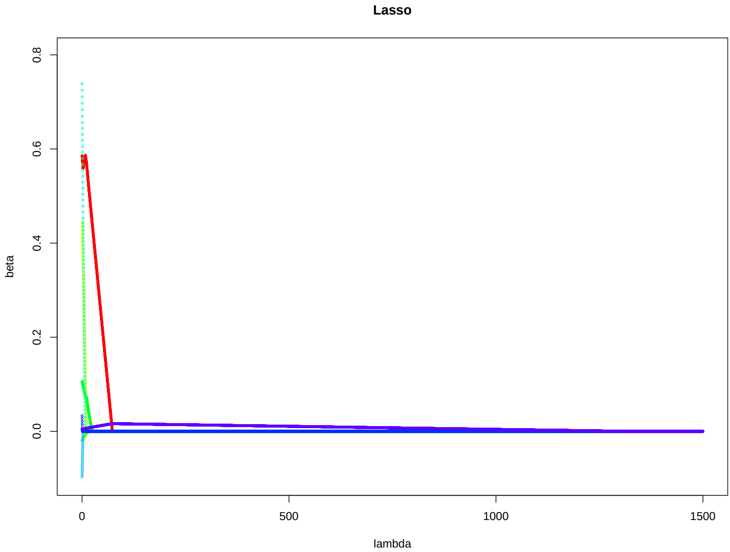
$$Y_i = \sum_{j=1}^{N} \beta_j X_{ij} + e_i \qquad i = 1, \ldots, m \qquad\qquad \min_{\beta} \sum_{i=1}^{m} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2$$

Making the model sparse. The lasso

$$\min_{\beta} \sum_{i=1}^{m} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2 + \lambda \|\beta\|_1$$

- R. Tibshirani, "Regression shrinkage and selection via the lasso", *J. of the Royal Statistical Society - B* , 1996
- $\approx$ 27.500 cites in Scholar


Lasso
beta
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
lambda


Lasso
beta
lambda
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500

$$\min_{\beta} \sum_{i=1}^{m} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2 + \lambda \|\beta\|_1$$

$$\min_{\beta} \sum_{i=1}^{m} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2 + \lambda \|\beta\|_1$$

- Lasso and relatives implemented in several packages in `R` (e.g. `lars`, `elasticnet`, …) and Python (`scikit-learn`)

$$\min_{\beta} \sum_{i=1}^{m} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2 + \lambda \|\beta\|_1$$

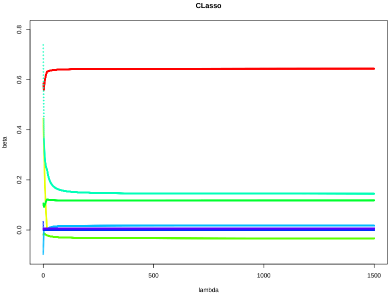
- Lasso and relatives implemented in several packages in `R` (e.g. `lars`, `elasticnet`, ...) and Python (`scikit-learn`)
- Records treated homogeneously. No control of errors on subpopulations, in case of heterogeneous data

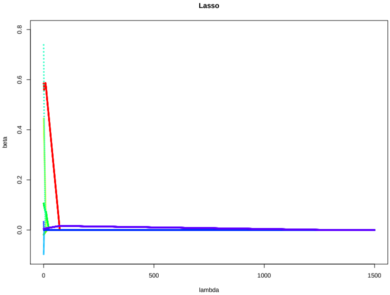
$$\min_{\beta} \sum_{i=1}^{m} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2 + \lambda \|\beta\|_1$$

- Lasso and relatives implemented in several packages in `R` (e.g. `lars`, `elasticnet`, ...) and Python (`scikit-learn`)
- Records treated homogeneously. No control of errors on subpopulations, in case of heterogeneous data
- New Mathematical Optimization problem:

$$\begin{array}{lll} \min_{\beta} & \sum_{i=1}^{m} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2 + \lambda \|\beta\|_1 & \\ \text{s.t.} & \sum_{i \in S_h} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2 \leq (1 + \tau_h) SSE_h & \forall h \end{array}$$

$$
\begin{aligned}
\min_{\beta} \quad & \sum_{i=1}^{m} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2 + \lambda \|\beta\|_1 \\
\text{s.t.} \quad & \sum_{i \in S_h} \left(Y_i - \sum_{j=1}^{N} \beta_j X_{ij} \right)^2 \leq (1 + \tau_h) SSE_h \qquad \forall h
\end{aligned}
$$





…but we don't know how to (easily) build the path

We invite members of ALIO and the worldwide Operations Research community to take part of XX Latin-Iberian-American Conference on Operations Research (CLAIO2020), to be held in Madrid (Spain), August 31st-September 2nd 2020. The conference is organized by the Latin-American Association of Operations Research Societies (ALIO), the Spanish Society of Statistics and Operations Research (SEIO), Universidad Complutense de Madrid (UCM) and Universidad Rey Juan Carlos (URJC). The academic program will consist of parallel, technical and special sessions, plenary talks and tutorials covering several aspects of OR.

Antonio Alonso-Ayuso (URJC), Javier Martín-Campo (UCM),

Conference Chairs of CLAIO 2020

Organizers

SEIO

Universidad Rey Juan Carlos

Universidad Complutense Madrid

Confirmed speakers:

Anna Nagurney. University of Massachusetts (USA)
Sebastian Ceria. Axioma (Argentina)
Emma Hart. University of Edimburg (UK)
Ángel Corberán. Universitat de Valencia (Spain)
Carlos Henggeler Antunes. Universidade de Coimbra (Portugal)

IFORS
INTERNATIONAL FEDERATION OF OPERATIONAL RESEARCH SOCIETIES

EURO
THE ASSOCIATION OF EUROPEAN OPERATIONAL RESEARCH SOCIETIES

APDIO
ASSOCIAÇÃO PORTUGUESA DE INVESTIGAÇÃO OPERACIONAL

ALIO

Many thanks!!!

ecarrizosa@us.es