Interpretable Credit Scoring. Methodological Adaptations to the Right to Explanation.

Belen Martin-Barragan
University of Edinburgh Business School
belen.martin@ed.ac.uk

Outline

- Why interpretable Credit Scoring?
- Background. Interpretable Support Vector Machines.
- Some encouraging results.
- Interpretable Support Vector Machines for functional data.
- Some encouraging results.
- EPSRC project: our plan.
- Conclusions

Why interpretable Credit Scoring?

Getting insight

Companies are still far from capturing the whole potential of data analytics. The biggest barrier is the struggle to incorporate data-driven insights into day-to-day processes.

McKinsey & Company, December 2016.

Why interpretable Credit Scoring?

Confidence

General public, regulators and practitioners are more confident with interpretable models than black-box type models.

Why interpretable Credit Scoring?

Right to an explanation

- General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679).
- Equal Credit Opportunity Act (Regulation B of the Code of Federal Regulations) in the US.

Individual rights to be given an explanation for decisions that significantly affect and individual, particularly legally or financially.

Why interpretable Credit Scoring?

Around 80% of the time devoted to data analysis is spent on pre-processing.
Original variables are transformed.
Transformations may improve performance, at the cost of interpretability.

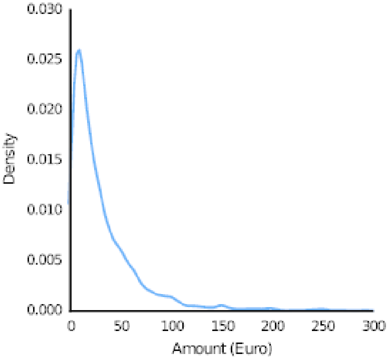
For example, in a credit card:

- Monthly balance.
- Do the owner spend more at the beginning of the month?
- Do the owner spend more at the weekends?
- Spending types: restaurants, transport, charities, ...
- Distribution of amounts incurred in transactions.
- ...

Why interpretable Credit Scoring?

Example

Distribution of amounts incurred in transactions.



Background. Interpretable Support Vector Machines.

Why Support vector machines?

- Considered in the **top-ten** algorithms in Data Mining
  - Wu et al. *Knowl. Inf. Syst.* (2008) 14:137
- Among the **most popular** keywords (Microsoft Academic Research Database):
  - 7th most popular in Data Mining
  - 4th most popular in Machine Learning and Pattern Recognition

Support Vector Machines

Interpretability:

- Nonlinear classifiers can be obtained using the kernels.
- Kernels can be seen as complex transformations.
- The kernel are difficult to interpret.
- Interpretation is a subjective issue.
- For example:
  - Is ROCE high?
  - Is EBIT Growth low?
- Visualizing the role of the predictor in the classifier.

Binarized Support Vector Machines

- We consider features of the form:

$$\phi_{jb}(x) = \begin{cases} 1, & \text{if } x_j \geq b \\ 0, & \text{otherwise.} \end{cases}$$

- The score function will be a combination of these features:

$$f(\mathbf{x}_i) = \beta_0 + \beta_{1b_1}\phi_{1,b_1}(x_i) + \ldots + \beta_{1b_1}\phi_{1,b_2}(x_i) + \ldots + \beta_{1b_1}\phi_{k,b_n}(x_i)$$

A problem arise: how to choose the thresholds?

Binarized Support Vector Machines

An example of data binarization

class	x_1	x_2	ϕ	ϕ	ϕ	ϕ	ϕ
-1	0.23	0.29	0	0	0	0	0
-1	0.72	0.32	1	1	0	0	0
-1	0.25	0.53	0	0	0	1	1
-1	0.52	0.05	0	0	0	0	0
1	0.61	0.43	1	0	0	1	0
1	0.92	0.95	1	1	1	1	1
1	0.76	0.47	1	1	1	1	0
			$x_1 \geq 0.61$?	$x_1 \geq 0.72$?	$x_1 \geq 0.76$?	$x_2 \geq 0.43$?	$x_2 \geq 0.53$?

Binarized SVM

IDEA

We theoretically consider all the possible thresholds.
We use $\|\cdot\|_1$ norm to regularize.

L_1-norm SVM:

$$
\begin{array}{lll}
\min & \|\boldsymbol{\beta}\|_1 + C\|\xi_i\|_1 & \\
\text{s.t.:} & \beta_0 + \boldsymbol{\beta}^\top \phi(\mathbf{x}_i) + \xi_i \geq 1, & \text{if } i \text{ is distressed} \\
& \beta_0 + \boldsymbol{\beta}^\top \phi(\mathbf{x}_i) + \xi_i \leq -1, & \text{if } i \text{ is healthy} \\
& \xi_i \geq 0 & \text{for each company } i.
\end{array}
$$

- :-) Sparse solution
- :-(Large problem
- :-) Formulated as a Linear Program !!! Solved via Column Generation

The application

Predicting default of a small business using different definitions of financial distress.
S-M Lin, J Ansell and G Andreeva, JORS 63, 539-548 (2012).

- A sample of UK SMEs (Basel II, turnover<€50M)
- Share price movements and financial statements (Datastream)
- 33 financial ratios where included, as in Lin et al.

Interpretability. Visualization tool.

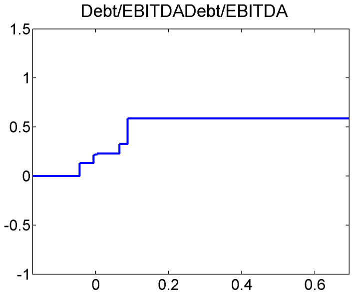
The role of predictor variable ℓ in the score function is modeled by the stepwise function

$$s \mapsto \sum_{\{b \in B_\ell \mid s \geq b\}} \beta_{\ell b}. \tag{1}$$

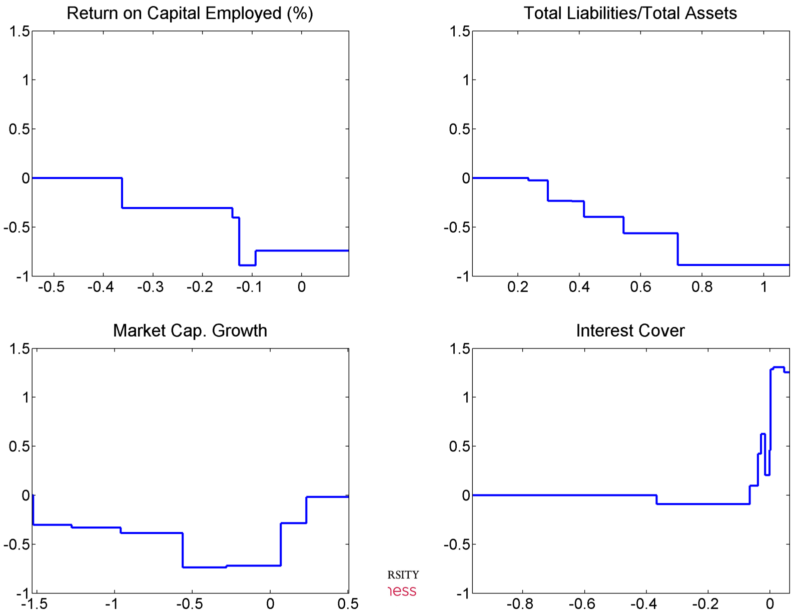
Interpretability. Visualization tool.

The role of predictor variable ℓ in the score function is modeled by the stepwise function

$$s \mapsto \sum_{\{b \in B_\ell \mid s \geq b\}} \beta_{\ell b}. \tag{1}$$



Interpretability. Results.



Binarized SVM. Results

- Performance measure: Area under the ROC Curve (AUC).
- Average from 10 fold cross validation.
- In each fold, an inner 10-fold cross validation was used to choose C.
- Missing values replaced by the mean.

Performance of the classifiers:

	BSVM	SVM
AUC	73.14	72.88

Other examples. Correct classification rates.

E. Carrizosa, B. Martin-Barragan, D. Romero Morales. Binarized Support Vector Machines, *INFORMS JoC* 22 (1), pp154-167, 2010.

	size	TreePr	TreeCr	SVM	BSVM
sonar	208 × 60	71.63	65.38	78.37	90.38
bands	277 × 56	62.82	67.51	71.12	70.40
credit	653 × 43	86.22	83.31	85.91	87.75
ionosphere	351 × 34	89.17	86.32	84.90	92.31
wdbc	569 × 30	92.09	92.62	96.66	97.01
cleveland	297 × 13	78.79	68.69	84.51	81.44
housing	506 × 13	83.99	84.78	85.38	86.97
pima	768 × 8	76.43	72.92	76.17	72.66
bupa	345 × 6	67.83	66.67	69.28	74.78

Table: Looc for BSVM and Benchmarking Methods

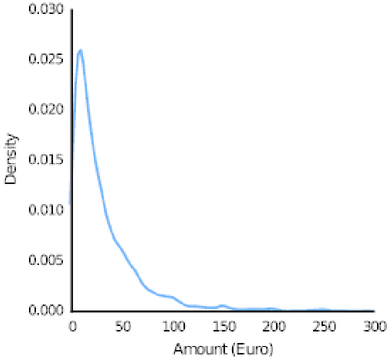
Interpretable SVM

- Interpretability:
  Insightful knowledge about the nonlinear behavior of the classifier.
- Prediction ability:
  Out-of-sample performance, better than CART & competitive vs SVM.
- Other advantages: robustness against outliers.
- Extensions:
  - when there are many variables, you may want to *prune* the classifier. CMR IJoC (2010)
  - interactions between variables. CMR EJOR (2011)
  - functional data analysis. MRL EJOR (2014)

Functional data analysis

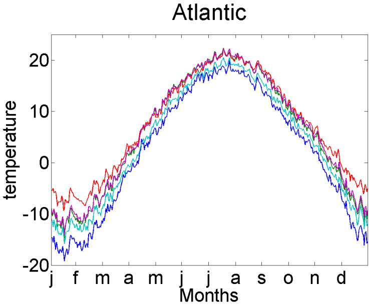
Example

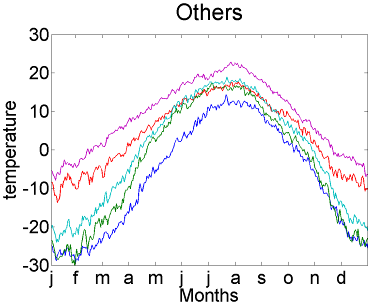
Distribution of amounts incurred in transactions.



Classification of Functional Data

FDA is the analysis of information on curves or functions.

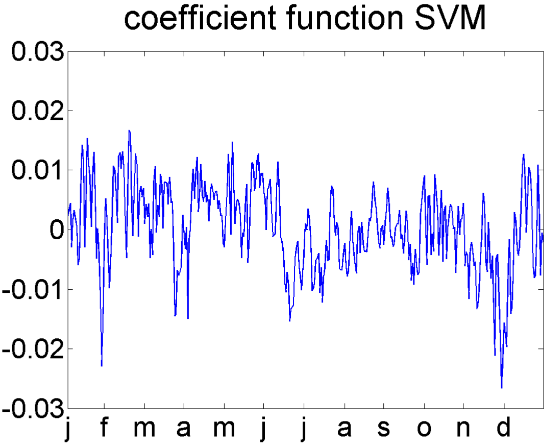




Weather data

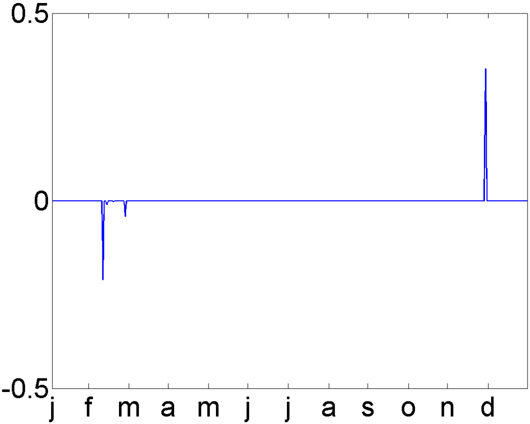
- One year of daily temperature measurements.
- 35 Canadian weather stations.
- Two classes: Atlantic climate vs. Others.

Interpretable Support Vector Machines for Functional Data

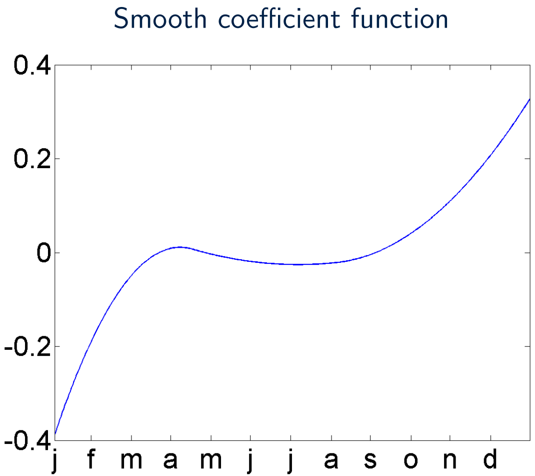


Interpretable Support Vector Machines for Functional Data

Detect relevant days

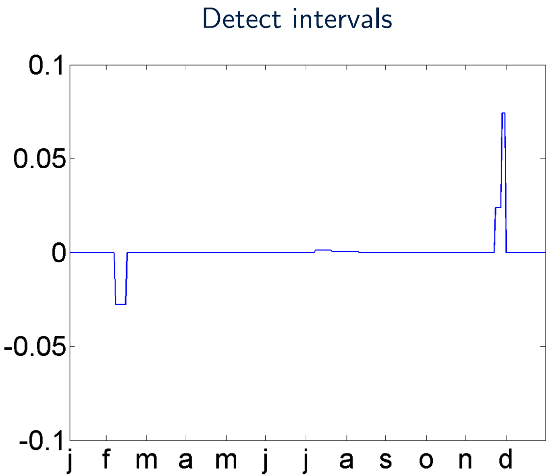


Interpretable Support Vector Machines for Functional Data



UNIVERSITY OF EDINBURGH Business School
CRC Credit Research Centre
EPSRC Engineering and Physical Sciences Research Council

Interpretable Support Vector Machines for Functional Data

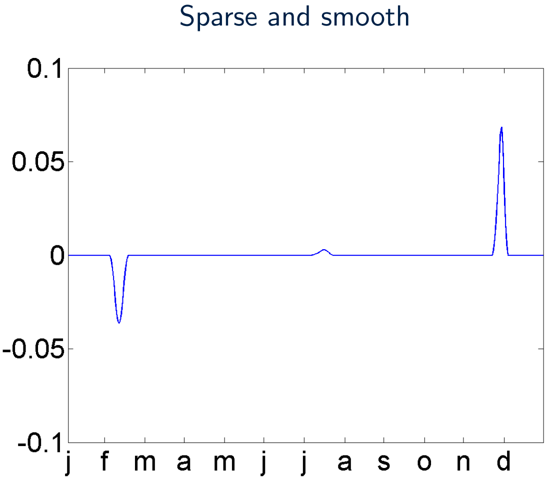


UNIVERSITY OF EDINBURGH Business School

CRC Credit Research Centre

EPSRC Engineering and Physical Sciences Research Council

Interpretable Support Vector Machines for Functional Data



UNIVERSITY OF EDINBURGH Business School

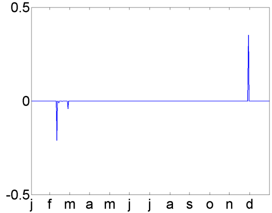
CRC Credit Research Centre

EPSRC Engineering and Physical Sciences Research Council

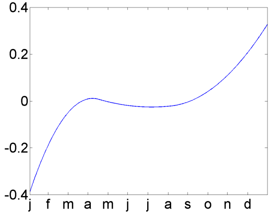
What about misclassification rates?

SVM applied to crude data: 5.7143

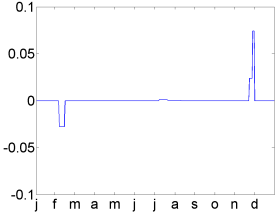
2.8571



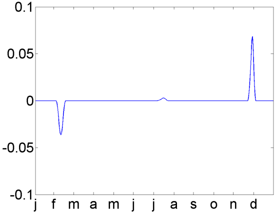
17.1429



5.7143



2.8571



EPSRC project: Optimization models for interpretable analytics.

Optimization models for interpretable analytics.

Developing new methodologies to estimate models that are easier to interpret than the current state-of-the-art methods.

- Probability of Default: Logistic Regression.
- Loss Given Default
  - Develop a new 1-phase method.
  - Combine ideas from SVM and Regression.
- Time to default: survival analysis.

EPSRC project: Optimization models for interpretable analytics.

Resources: 120k pounds

- Workshop 1: Right to an explanation. What does it mean?
- Workshop 2: Validation workshop.
- Postdoc position.
- Dissemination of results.

Conclusions.

Take-home messages

- A need for advanced algorithms that produce interpretable results.
- Interpretability is subjective. Explore definitions and model them.
- False trade-off interpretability and prediction ability.

Keep in touch

I you are interested ...

belen.martin@ed.ac.uk.