Adaptative splines-based logistic regression with a ReLU neural network

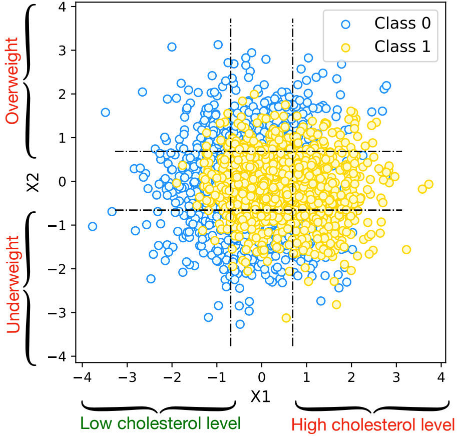
Marie Guyomard, Susana Barbosa, Lionel Fillatre

JOBIM 2022

Sommaire

Contextualization

Simulated Data



Classes :

 : healthy

 : sick

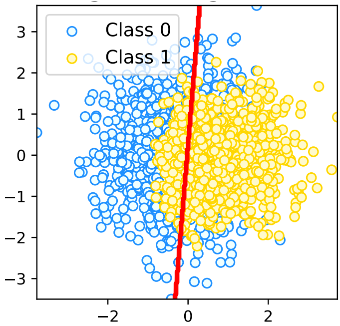
Features :

$X1$: cholesterol

$X2$: weight

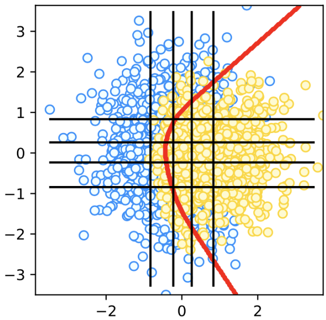
Linear vs non-linear Classification

Logistic Regression Linear



Global accuracy : 70%

Logistic Regression Splines



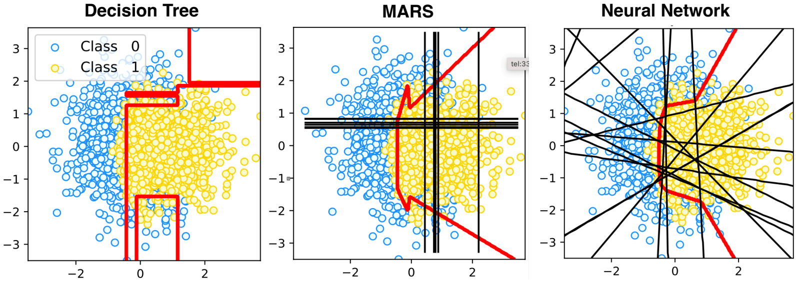
Global accuracy : 76%

Legend : estimated boundary decision in red, segmentation in black

Non-linear Classification

Main difficulty : estimate simultaneously tresholds and the decision rule.

State of the art :



Decision Tree

- [x] Classification task
- [x] Partition with hyperplans
- [x] Interpretabilty
- [ ] Optimization

MARS

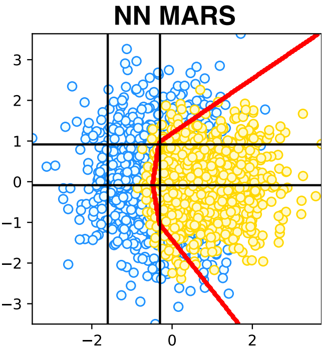
- [ ] Classification task
- [x] Partition with hyperplans
- [x] Interpretabilty
- [ ] Optimization

Neural Network

- [x] Classification task
- [ ] Partition with hyperplans
- [ ] Interpretabilty
- [x] Optimization

Objectives

- Classification task
- Modeling of non-linear phenomena : threshold effects
- Global optimization
- Interpretability



☑ Classification task
☑ Partition with hyperplans
☑ Interpretabilty
☑ Optimization

Problem Statement

Notations

Let suppose we have N independent and identically distributed pairs $(x^{(i)}, y^{(i)})$ with

- $x^{(i)} \in \mathbb{R}^d$: explanatory variables
- $y^{(i)} \in \{0, 1\}$: binary label

Binary Classification Task

Bayesian Maximum a Posteriory Classifier

$$\delta : \mathbb{R}^d \longrightarrow [0,1]$$
$$\delta(x) = \underset{k=\{0,1\}}{\arg\max}\, \hat{\mathbb{P}}_\theta \left(Y = k | X = x \right). \quad (1)$$

Logistic Regression

$$\hat{\mathbb{P}} \left(Y = 1 | X = x \right) = \sigma \left(\psi(x) \right) = \frac{1}{1 + \exp\left(-\psi(x) \right)}. \quad (2)$$

Logistic Regression

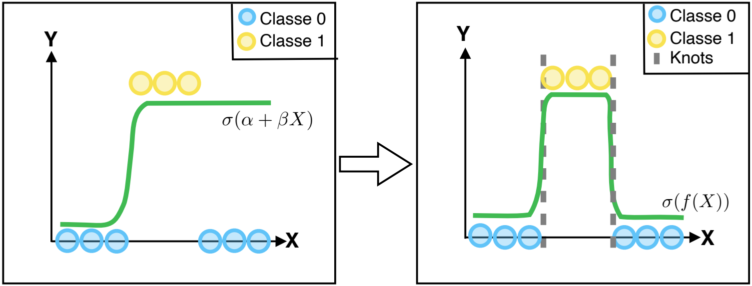


Figure – Logistic Regression classifiers : linear (left) and non-linear (right).

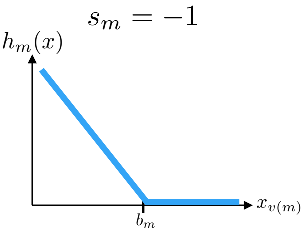
How can we model the non-linear effects in the score function ?

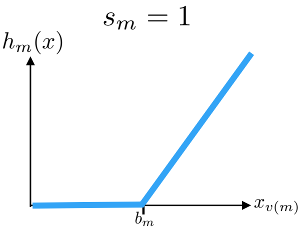
MARS

$$\psi^{\text{MARS}}(x) = \sum_{m=1}^{M} \beta_m h_m(x), \tag{3}$$

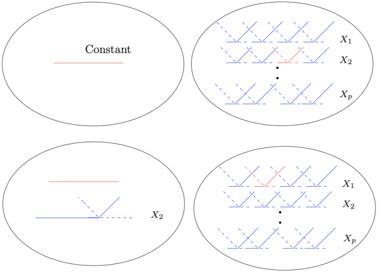
$$h_m(x) = [s_m(x_{v(m)} - b_m)]_+ \tag{4}$$

$$= \begin{cases} \max\{0, x_{v(m)} - b_m\}, & \text{if } s_m = 1, \\ \max\{0, b_m - x_{v(m)}\}, & \text{if } s_m = -1. \end{cases} \tag{5}$$





Main limit of MARS



Source : Hastie & Tibshirani, 2009.

Greedy Optimization :

- ✗ Global optimality
- ✗ Uncontrolable segmentation
- ✗ Over-segmentation

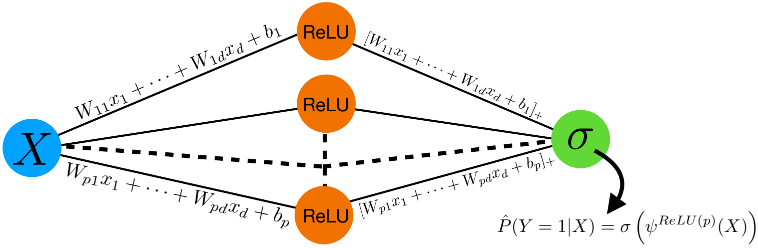
How can we model a non linear logistic regression while using a global criterion ?

ReLU Neural Network

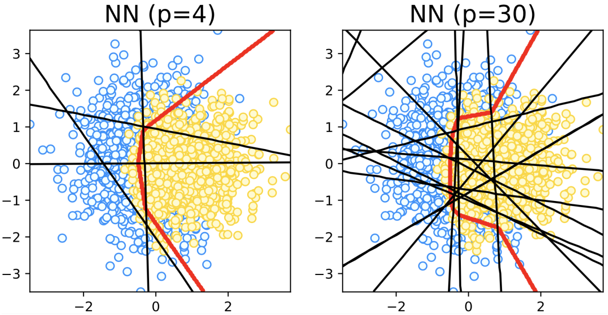
$$\Phi^{(p)}: \quad x \longrightarrow \hat{y}(x)$$
$$x \longrightarrow \sigma \circ \psi^{\text{ReLU(p)}}. \tag{6}$$

with σ the sigmoid defined before, $\beta \in \mathbb{R}^p$, $W \in \mathbb{R}^{p \times d}$ et $b \in \mathbb{R}^p$.

$$\psi^{\text{ReLU(p)}}(x) = \beta_0 + \beta^T \left[Wx + b\right]_+ . \tag{7}$$



- NN can approximate splines [Balestriero]
- NN can approximate MARS [Eckle]
  ✔ theory ✘ pratical
- NN makes a partition with oblique regions
  ✘ 'Black Box'



Neural Network MARS

NN-MARS

Main idea : couple the Logistic Regression with the MARS model while using Neural Network.

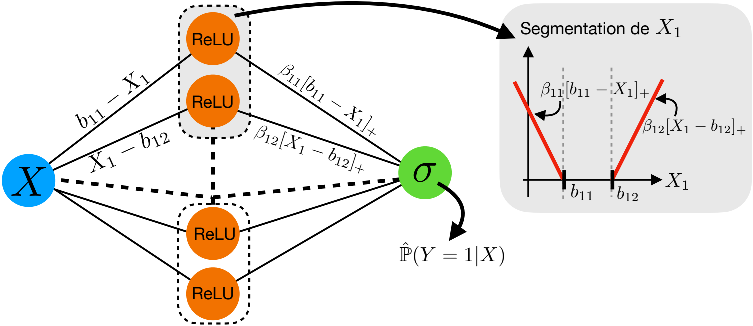
The proposed method :

- Controles & automatizes the segmentation of variables
  - We know from doctors' feedback that over-segmenting a biological variable is not relevant
  - Hyperplans
- Interpretability
- Easy to use for doctors

NN-MARS

$$\psi^{\text{NN-MARS}}(x) = \beta_0 + \sum_{j=1}^{d} g_j(x_j), \tag{8}$$

$$g_j(t) = \beta_{j1}[b_{j1} - t]_+ + \beta_{j2}[t - b_{j2}]_+, \quad t \in \mathbb{R}. \tag{9}$$



Experiments

Experiments on real a data set

Parkinson database

- Predict Parkinson from voice recordings
- 195 patients (24,6% with Parkinson)
- We kept d = 16 biomedical recordings

Tested Methods

- NN-MARS
- Logistic Regression
- Logistic Regression Natural Cubic Splines
- Decision Tree
- MARS
- NN ReLU

Predictive accuracies

	Training		Test	
	Accuracy	AUC	Accuracy	AUC
LR	85 (2)	87 (2)	76 (1)	80 (6)
DT	91 (2)	94 (2)	88 (1)	77 (3)
LR SCN	90 (2)	94 (1)	82 (3)	87 (5)
MARS	90 (3)	91 (6)	82 (4)	89 (4)
NN ($p = 16$)	87 (4)	91 (7)	81 (6)	88 (7)
NN ($p = 70$)	86 (4)	91 (7)	83 (6)	88 (7)
NN-MARS	87 (1)	92 (3)	83 (5)	91 (5)

Table – Mean in % and (standard-deviation) of predictive accuracies obtained after a 5-folds cross-validation : DT, LR NCS, NN with 16 neurons, NN with à 70 neurons & the NN-MARS.

Interpretability

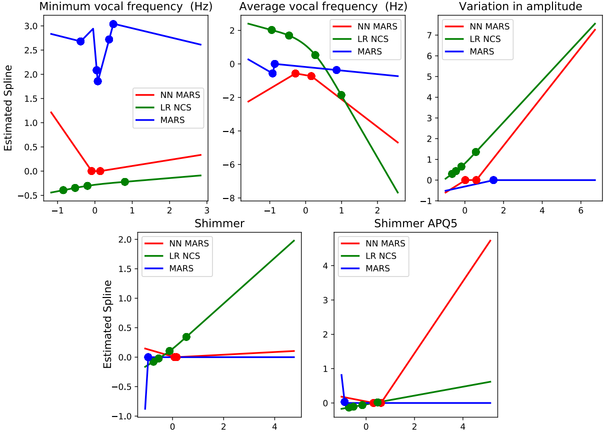


Figure – Estimated splines for the Minimum voice frequency, Average voice frequency, Amplitude variation, Shimmer, Shimmer APQ5.

Conclusion & Future Works

Conclusion & Future Works

Conclusion

- Binary Classification
- Automatized & controled discretization of the variables
- Powerfull & Interpretable
- Biologicaly relevant & easy use

Future Works

- Categorical data
- Interactions between the variables

Bibliographie

- Douglas M Hawkins, On the choice of segments in piecewise approximation, *IMA Journal of Applied Mathematics*, 9(2) :250–256, 1972.
- Asher Tishler and Israel Zang, A new maximum likelihood algorithm for piecewise regression, *Journal of the American Statistical Association*, 76(376) :980–987, 1981.
- Konstantin Eckle et al, A comparison of deep networks with relu activation function and linear spline-type methods, *Neural Networks*, 110 :232–242, 2019.
- Jerome H Friedman, Multivariate adaptive regression splines, *The annals of statistics*, 19(1) :1–67, 1991.
- Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, *MIT Press*, Cambridge, MA, USA, 2016.
- Randall Balestriero et al, A spline theory of deep learning, *In International Conference on Machine Learning*, pages 374–383. PMLR, 2018.
- Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman, The elements of statistical learning : data mining, inference, and prediction, volume 2, *Springer*, 2009.
- Xiao-Hua Zhou, Donna K McClish, and Nancy A Obuchowski, Statistical methods in diagnostic medicine, *John Wiley Sons*, 2009.
- Max Little et al, Suitability of dysphonia measurements for telemonitoring of parkinson's disease, *Nature Precedings*, 2008.