Adaptative spline-based Logistic Regression with a ReLU Neural Network

GRETSI 2022

Marie Guyomard[1], Susana Barbosa[2], Lionel Fillatre[1]

[1] Université Côte d'Azur, CNRS, I3S

[2] Université Côte d'Azur, CNRS, IPMC

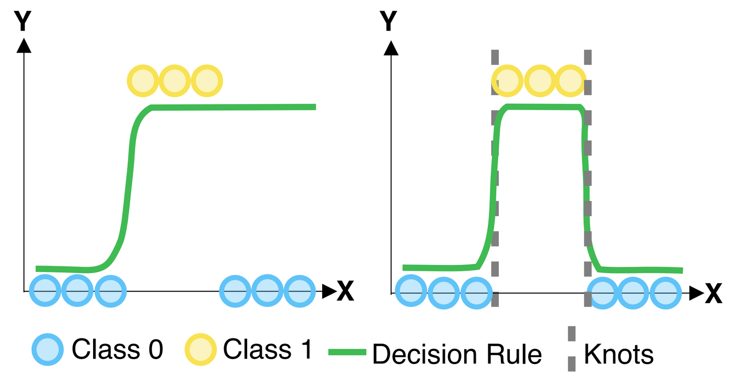
Contact: `guyomard@i3s.unice.fr`

06–09/09/2022

This work has been supported by the French government, through the UCA DS4H Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-17-EURE-0004.

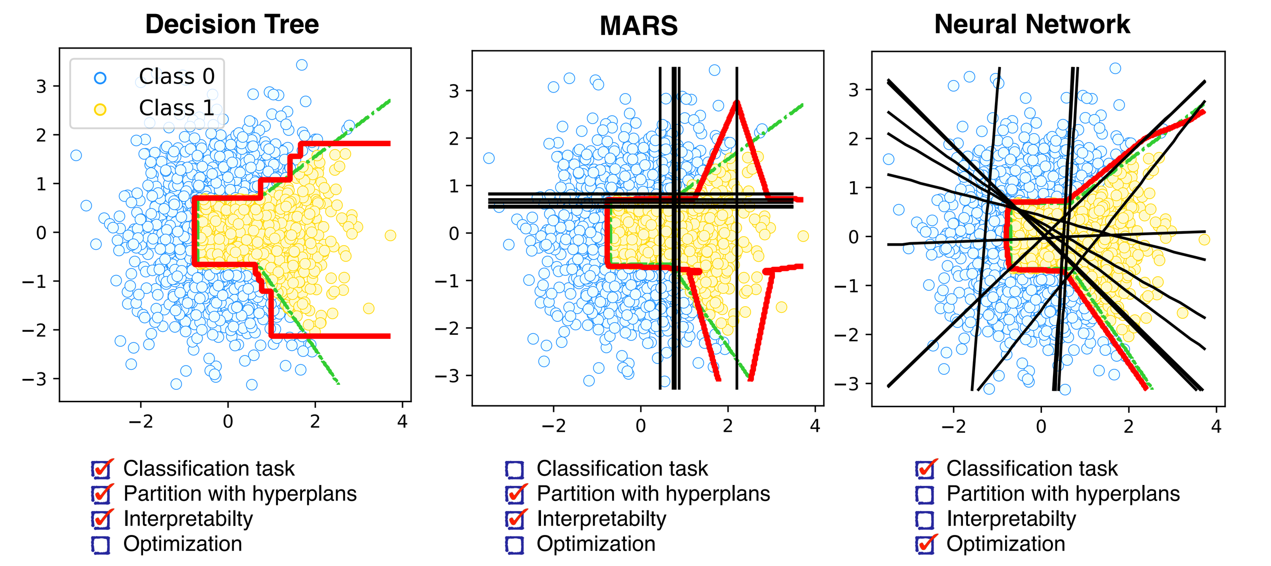
1 INTRODUCTION

Objectives



- Classification task
- Modeling of non linear effects
- Discretization of variables
- Interpretability

State of the art



2 PROBLEM STATEMENT

Binary Classification Task

We have N independent and identically distributed realizations $(x^{(i)}, y^{(i)})$ of the couple (X, Y) where $X \in \mathbb{R}^d$ is the feature vector and $Y \in \{0,1\}$ is the label:

$$Y = f(X)$$

.

Logistic Regression

$$\mathbb{P}(Y=1|X=x) = \sigma(\psi(x)) = \frac{1}{1+\exp(-\psi(x))}, \text{ with } \psi(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_d x_d. \quad (1)$$

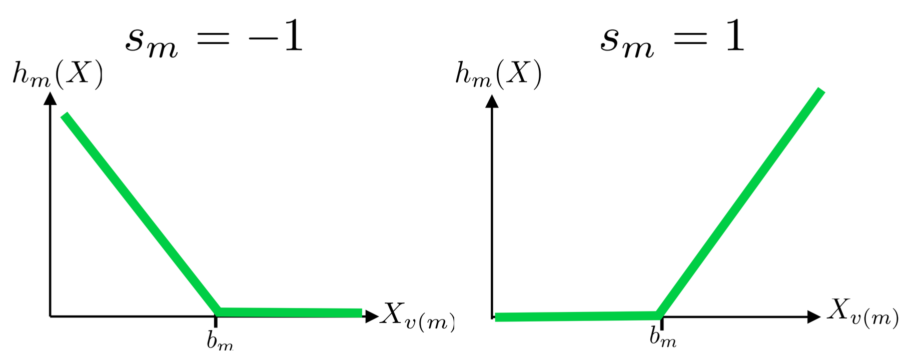
MARS Modeling

$$\psi^{\text{MARS}}(x) = \sum_{m=1}^{M} \beta_m h_m(x), \quad (2)$$

where $h_m(x)$ is a spline function of the form

$$h_m(x) = [s_m(x_{v(m)} - b_m)]_+ \quad (3)$$

$$= \begin{cases} \max\{0, x_{v(m)} - b_m\}, & \text{if } s_m = 1, \\ \max\{0, b_m - x_{v(m)}\}, & \text{if } s_m = -1. \end{cases} \quad (4)$$



Modeling of $h_m(x)$.

3 NEURAL NETWORK NN-MARS

Ideas

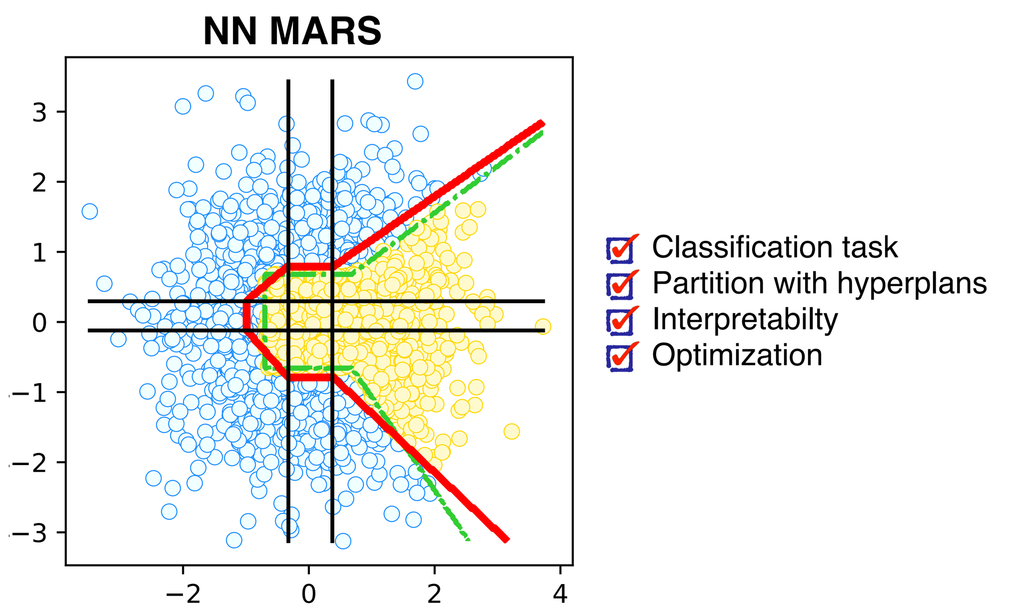
Develop a NN inspired by the MARS Model

We know that :

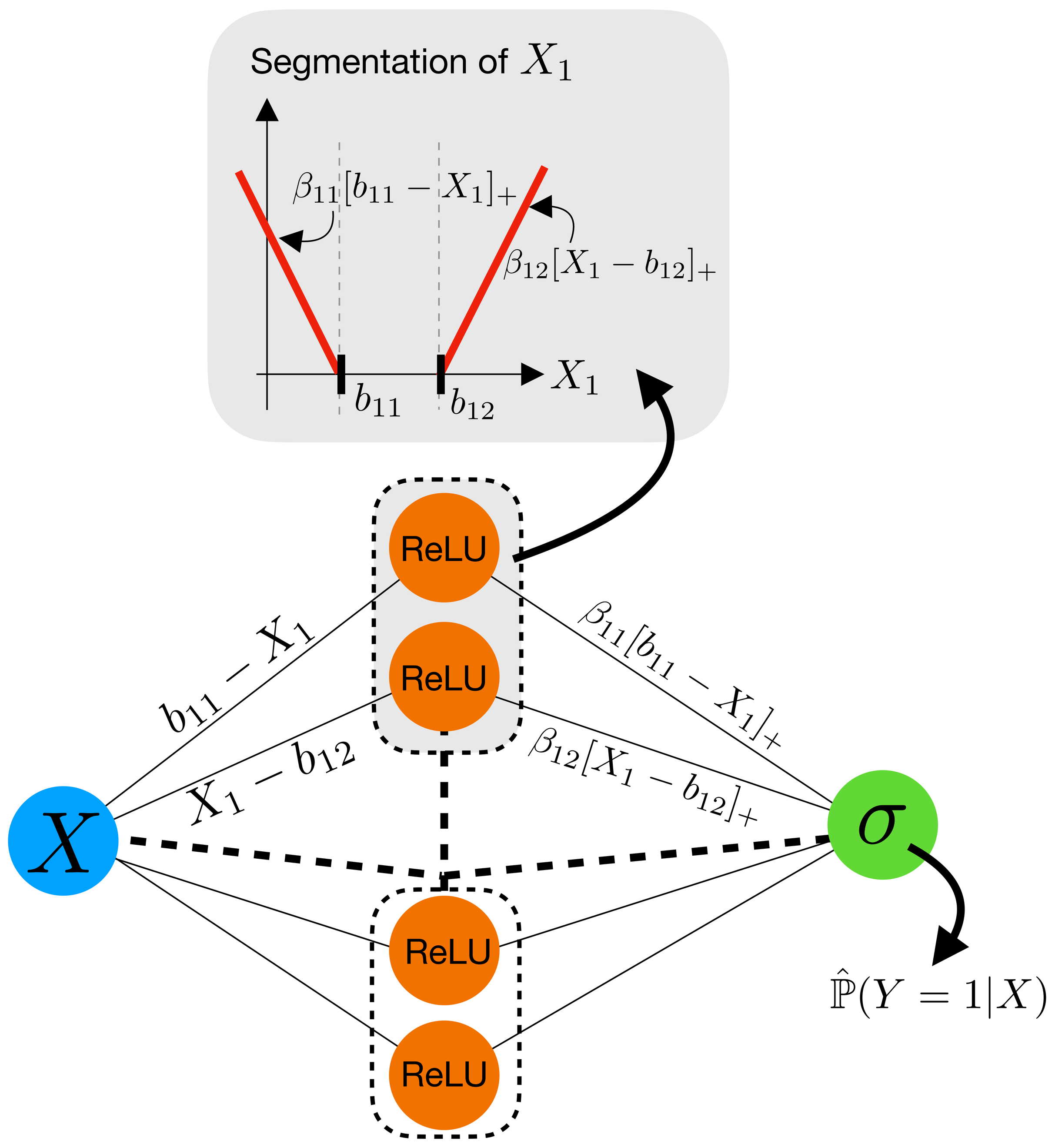
- NN can approximate splines [Balestriero]
- NN can approximate MARS [Eckle]
  ✔ theory ✘ pratical
- NN makes a partition with oblique regions
  ✘ 'Black Box'

The proposed method **NN-MARS** :

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



Modeling



Architecture of NN-MARS: the inputs, the hidden layer and the estimated labels.

ReLU Neural Network

Let Φ be a ReLU Neural Network for classification with a layer of p neurons :

$$\Phi^{(p)} : \begin{array}{l} x \longrightarrow \hat{y}(x) \\ x \longrightarrow \sigma \circ \psi^{\text{ReLU(p)}}. \end{array} \quad (5)$$

with σ defined by (1), $\beta \in \mathbb{R}^p$, $W \in \mathbb{R}^{p \times d}$ and $b \in \mathbb{R}^p$.

$$\psi^{\text{ReLU(p)}}(x) = \beta_0 + \sum_{i=1}^{p} \beta_i \left[\sum_{j=1}^{d} W_{i,j} x_j + b_i \right]_+. \quad (6)$$

with $[\cdot]_+$ defined by the equation (4).

NN-MARS

The NN-MARS is constructed from this model, with constraints on the weights W :

$$\psi^{\text{NN-MARS}}(x) = \beta_0 + \sum_{j=1}^{d} g_j(x_j), \quad (7)$$

$$g_j(t) = \beta_{j1}[b_{j1} - t]_+ + \beta_{j2}[t - b_{j2}]_+, \quad t \in \mathbb{R}. \quad (8)$$

4 EXPERIMENTS

Data

Parkinson database :

- Predict Parkinson from voice recordings
- We kept $d = 16$ biomedical recordings
- 5-folds Cross Validation

Tested Methods

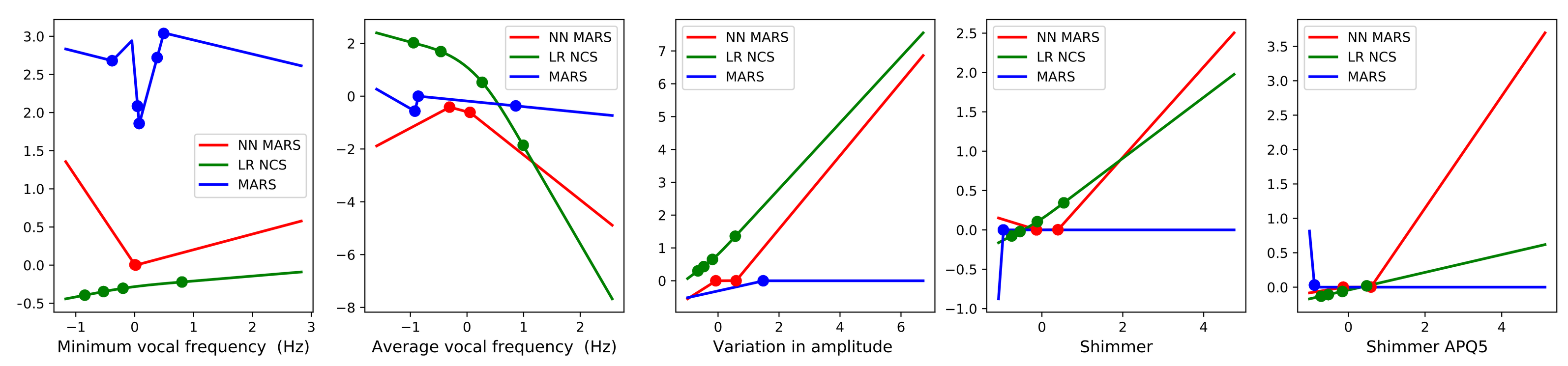
We compare the performance and the explicability of :

- NN-MARS (7)
- Logistic Regression (1)
- Decision Tree [Hastie - Section 9.2]
- Logistic Regression Natural Cubic Splines [Hastie - Section 5.2]
- MARS (2)
- ReLU NN (5)

Results

	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)**

- Importance of non linear effects (LR less efficient than non linear methods)
- NN-MARS obtains the best AUC on the testing sample
  - automatization of knots (✘ LR NCS)
  - control of the segmentation
- NN-MARS more stable that ReLU NNs
- MARS : only 5 variables segmented
  ✘ greedy learning



5 CONCLUSION

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

6 FUTURE WORKS

+ Categorical data
+ Interactions between the variables

7 REFERENCES

- Konstantin Eckle et al. A comparison of deep networks with relu activation function and linear spline-type methods. *Neural Networks*, 110:232–242, 2019.
- 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.