Kernel Logistic Regression Approximation of an Understandable ReLU Neural Network

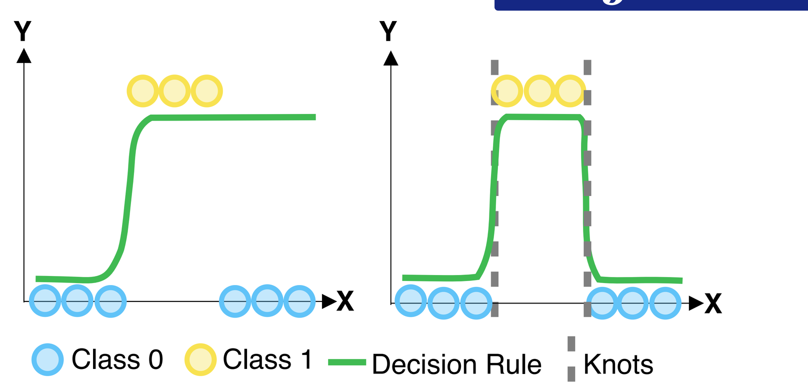
Marie Guyomard[1], Susana Barbosa[2], Lionel Fillatre[1]
[1] Université Côte d'Azur, CNRS, I3S, [2] Université Côte d'Azur, CNRS, IPMC

Contacts: marie.guyomard83@gmail.com, lionel.fillatre@i3s.unice.fr

ICML International Conference On Machine Learning
Honolulu, Hawaii
23-29 July 2023

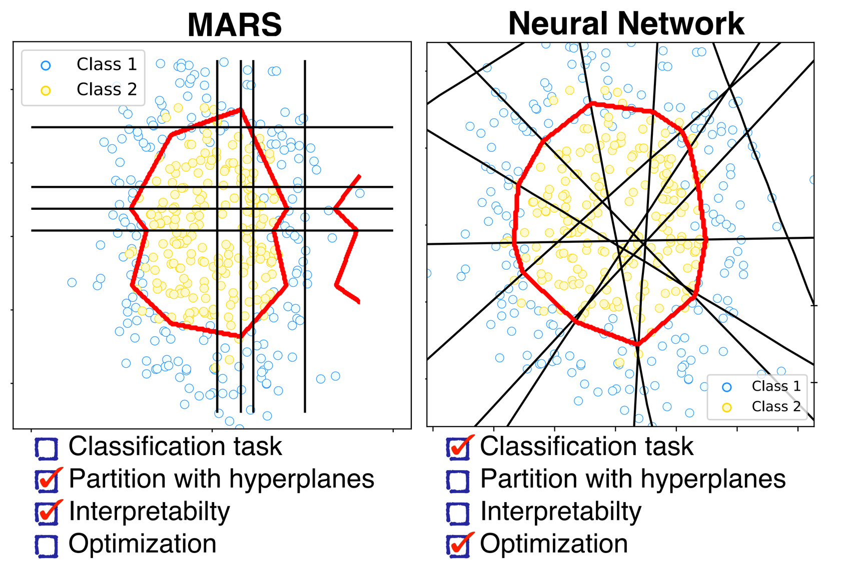
1 INTRODUCTION

Objectives

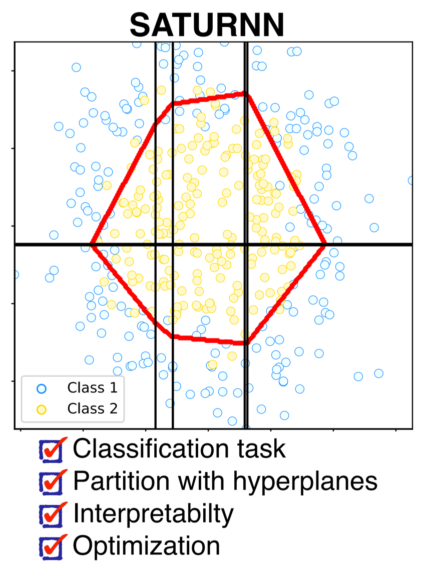


- Binary Classification task
  - Non linear effects
  - Discretization
- Interpretability
- Convergence guarantees

State of the art



Contribution



2 PROBLEM STATEMENT

Logistic Regression

$$\mathbb{P}(Y=1|X=x) = \sigma(\psi(x)) = \frac{1}{1+\exp(-\psi(x))},$$

with $\psi(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_d x_d$.

ReLU NN

Let $\delta_{\text{ReLU}}(x)$ be a ReLU Neural Network for binary classification with a layer of p neurons:

$$\delta_{\text{ReLU}}(x) = \sigma \circ \psi^{\text{ReLU}} = \sigma\left(\beta_0 + \sum_{i=1}^{p} \beta_i \phi\left(\sum_{j=1}^{d} \mathbf{W_{i,j}} x_j + b_i\right)\right),$$

with $\phi(\cdot) = \max(0,\cdot)$ the ReLU function and σ the sigmoid.

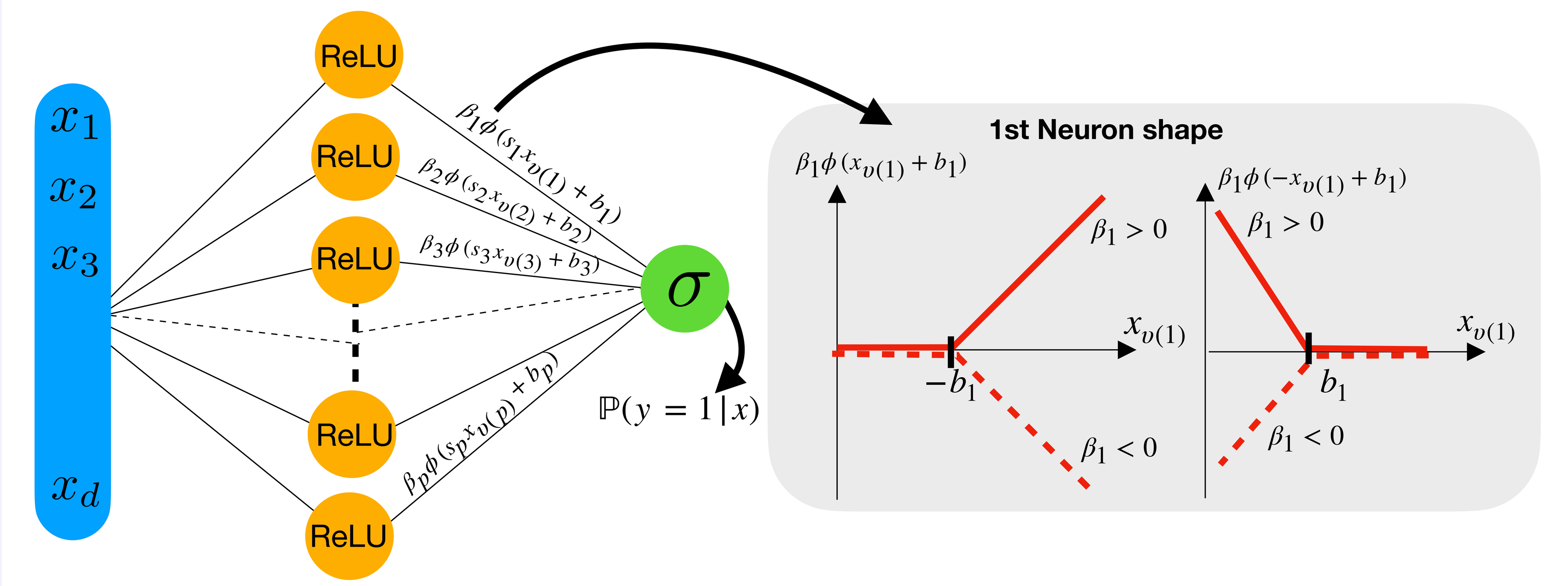
Black Box Model

- No interpretability
  NNs make a partition of the input space [Balestriero, 2018] with oblique regions : $\mathbf{W} \in \mathbb{R}^{p\times d}$ is a blender matrix.
- No guarantees on the convergence

Motivations

- ReLU NN can approximate MARS [Eckle, 2019] and GAM [Agarwal, 2021]
- NNs may asymptotically become linear with respect to their parameters as NN width p increases [Jacot, 2018]
  - ✗ Gaussian initializations [Lee, 2017]
  - ✗ Linear output layer [Liu, 2020]

3 SATURNN



Modeling

The SATURNN is constructed as a ReLU NN with constraints on the weight W:

$$\delta_{\text{SATURNN}}(x) = \sigma \circ \psi(x,\theta),$$
$$\psi(x,\theta) = \frac{1}{\sqrt{p}}\left[\beta_0 + \sum_{k=1}^{p} \beta_k \phi(s_k x_{v(k)} + b_k)\right],$$

- an input selector: $v(k) \sim \mathcal{U}[1,\dots,d]$
- an indicator of the shape: $s_k = \{-1,1\} \sim \mathcal{B}(1/2)$

The SATURNN is a special case of **GAMs**:

$$\psi(x,\theta) = \frac{1}{\sqrt{p}}\left[\beta_0 + \sum_{i=1}^{d} f_i(x_i)\right],$$
$$\text{with } f_i(x_i) = \sum_{1\le k\le p: v(k)=i} \beta_k \phi(s_k x_i + b_k).$$

Assumptions

Sample

$$\{(x^{(i)}, y^{(i)})\}_{i=1}^{N},\ x^{(i)} \in \mathbb{R}^d : \|x^{(i)}\|_2 \le r$$

Initializations

$$\theta^{(0)} = [\beta_0^{(0)},\dots,\beta_p^{(0)}, b_1^{(0)},\dots b_p^{(0)}]$$
$$\beta_k^{(0)} \sim \mathcal{N}(0,1), \qquad b_k^{(0)} \sim \mathcal{U}[-r,r]$$

Learning

$$\|\hat{\theta} - \theta^{(0)}\|_2 \le R,$$

- $R > 0$
- $\hat{\theta}$ the estimated parameters

Kernel Approximation

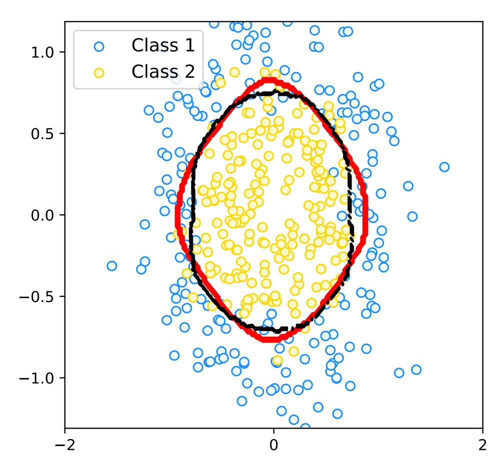
Approximation of SATURNN with Kernel Logistic Regression (KLR)

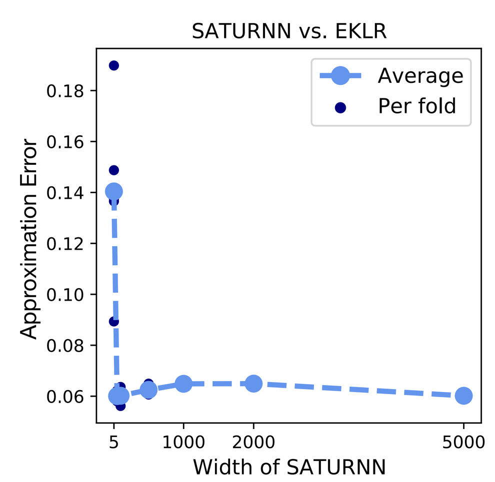
$$\delta_{KLR}(x,\alpha) = \sigma\left(\sum_{i=1}^{N} \alpha_i \kappa_0\left(\mathbf{x^{(i)}},\mathbf{x}\right)\right),$$

$$\kappa_0(x,\tilde{x}) = \frac{1}{p}\Bigg[1 + \sum_{k=1}^{p} \phi\left(s_k x_{v(k)} + b_k^{(0)}\right)\phi\left(s_k \tilde{x}_{v(k)} + b_k^{(0)}\right) + \beta_k^{(0)2}\mathbb{1}_{\{s_k x_{v(k)} + b_k^{(0)} > 0\}}\mathbb{1}_{\{s_k \tilde{x}_{v(k)} + b_k^{(0)} > 0\}}\Bigg].$$

Approximation of SATURNN with Expected KLR (EKLR)

$$\kappa(x,\tilde{x}) = \mathbb{E}(\kappa_0(x,\tilde{x})) = \frac{1}{p} + \frac{r^2}{6} + \frac{1}{4rd}\sum_{i=1}^{d}\left(2r(x_i\tilde{x}_i + 1) - |x_i - \tilde{x}_i| + \frac{1}{6}|x_i - \tilde{x}_i|^3\right).$$





4 EXPERIMENTS

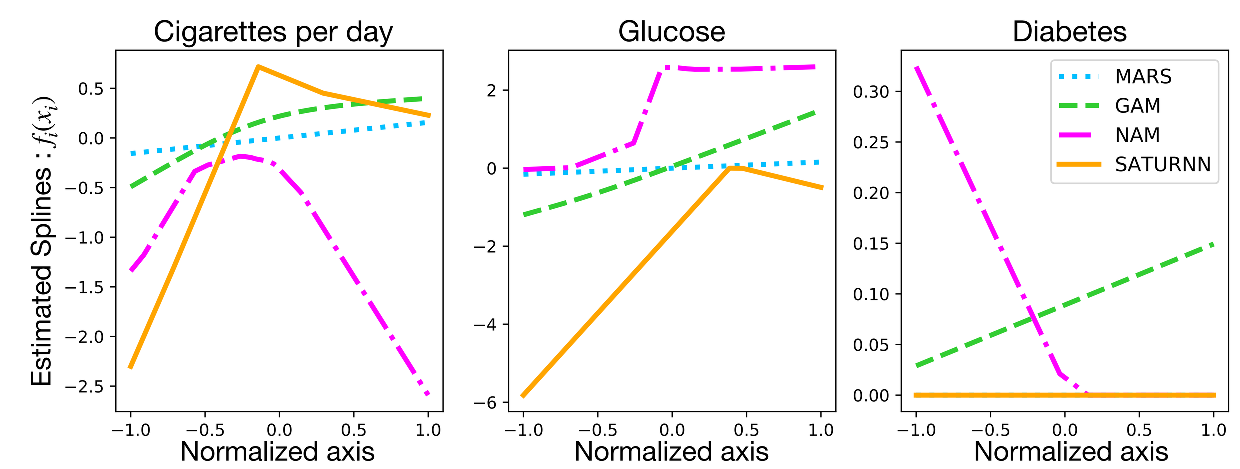
Data

Framingham dataset [Mahmood et al., 2014]

- 15 variables
- 1114 samples
- Prediction of a cardiovascular event (True / False)

Results

Methods	AUC Train	AUC Test	Computation Time
RF	0.76 (0.01)	0.70 (0.02)	0.1
MARS	0.75 (0.01)	0.71 (0.01)	0.1
GAM	0.8 (0.01)	0.69 (0.02)	0.65
EBM	0.77 (0.01)	0.72 (0.01)	7.8
NAM	0.78 (0.01)	0.70 (0.02)	694
RN ReLU	0.85 (0.02)	0.66 (0.03)	483
SATURNN	0.74 (0.02)	0.72 (0.02)	735
SATURNN_∞	0.84 (0.01)	0.69 (0.02)	591
KLR	0.74 (0.01)	**0.73 (0.02)**	0.32
EKLR	0.74 (0.01)	**0.73 (0.02)**	0.35



5 REFERENCES

- Balestriero, R. et al. A spline theory of deep learning. *In International Conference on Machine Learning*, 2018.
- Eckle, K. et al. A comparison of deep networks with relu activation function and linear spline-type methods. *Neural Networks*, 110:232–242, 2019.
- Agarwal, R. et al. Neural additive models: Interpretable machine learning with neural nets. *Advances in Neural Information Processing Systems*, 34: 4699–4711, 2021.
- Jacot, A., et al. Neural tangent kernel: Convergence and generalization in neural networks. *Advances in neural information processing systems*, 31, 2018.
- Lee, J.et al. Deep neural networks as gaussian processes. *arXiv preprint*, 2017.
- Liu, C., and al. On the linearity of large non-linear models: when and why the tangent kernel is constant. *Advances in Neural Information Processing Systems*, 33:15954–15964, 2020a.
- Mahmood, S. S., Levy, D., Vasan, R. S., and Wang, T. J. The framingham heart study and the epidemiology of cardiovascular disease: a historical perspective. *The lancet*, 383(9921):999–1008, 2014.

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.