Aalto-yliopisto
Perustieteiden
korkeakoulu

P-split formulation for neural networks

Venla Kjäll

23.05.2024

Ohjaaja: *Yu Liu*

Valvoja: Fabricio Oliveira

Systeemianalyysin laboratorio

Table of contents

Background – ReLU networks as surrogate models

- The hidden layers use only ReLU activation functions
  - $ReLU(y) := \max\{0, y\}$ (componentwise)
- Output layer uses the identity activation
- The models are not suitable for training, they are useful for finding optimal input examples for a given trained neural network

Background – 0-1 MILP formulation

$$\min \sum_{k=0}^{K} \sum_{j=1}^{n_k} c_j^k x_j^k + \sum_{k=1}^{K} \sum_{j=1}^{n_k} \gamma_j^k z_j^k$$

$$\left.\begin{aligned} \sum_{i=1}^{n_{k-1}} w_{ij}^{k-1} x_i^{k-1} + b_j^{k-1} &= x_j^k - s_j^k \\ x_j^k, s_j^k &\geq 0 \\ z_j^k &\in \{0, 1\} \\ z_j^k = 1 \rightarrow x_j^k &\leq 0 \\ z_j^k = 0 \rightarrow s_j^k &\leq 0 \end{aligned}\right\} k = 1, \ldots, K, \;\; j = 1, \ldots, n_k$$

$$lb_j^0 \leq x_j^0 \leq ub_j^0, \qquad j = 1, \ldots, n_0$$

$$\left.\begin{aligned} lb_j^k \leq x_j^k \leq ub_j^k \\ \overline{lb}_j^k \leq s_j^k \leq \overline{ub}_j^k \end{aligned}\right\} k = 1, \ldots, K, \;\; j = 1, \ldots, n_k.$$

Matteo Fischetti, Jason Jo, (2018). Deep neural networks and mixed integer linear optimization

Background – P-split formulation

- ReLU activation function can also be represented as a disjunction: $\begin{bmatrix} y = w^T x + b \\ w^T x + b \geq 0 \end{bmatrix} \vee \begin{bmatrix} y \quad = 0 \\ w^T x + b \leq 0 \end{bmatrix}$
- We partition the variables into P sets and split the constraints in the disjuction into P parts
- Should be computationally light but provide a tight relaxation
- The 1-split formulation is equal to the big-M formulation

Background – P-split formulation

$$\alpha_s^l = \sum_{d \in \mathcal{D}} \nu_d^{\alpha_s^l} \qquad \forall\, s \in \{1, \ldots, P\},\ \forall\, l \in \mathcal{D}$$

$$\sum_{s=1}^{P} \nu_l^{\alpha_s^l} \leq b_l \lambda_l \qquad \forall\, l \in \mathcal{D}$$

$$\underline{\alpha}_s^l \lambda_d \leq \nu_d^{\alpha_s^l} \leq \bar{\alpha}_s^l \lambda_d \qquad \forall\, s \in \{1, \ldots, P\}, \forall\, l, d \in \mathcal{D}$$

$$\sum_{i \in \mathcal{I}_s} h_{i,l}(x_i) \leq \alpha_s^l \qquad \forall\, s \in \{1, \ldots, P\},\ \forall\, l \in \mathcal{D}$$

$$\sum_{l \in \mathcal{D}} \lambda_l = 1, \quad \boldsymbol{\lambda} \in \{0, 1\}^{|\mathcal{D}|}$$

$$\boldsymbol{x} \in \mathcal{X}, \boldsymbol{\alpha}^l \in \mathbb{R}^P,\ \boldsymbol{\nu}^{\alpha_s^l} \in \mathbb{R}^{|\mathcal{D}|} \quad \forall\, s \in \{1, \ldots, P\},\ \forall\, l \in \mathcal{D}\ ,$$

Jan Kronqvist, Ruth Misener, Calvin Tsay, (2022). P-split formulations: A class of intermediate formulations between big-M and convex hull for disjunctive constraints

Objective

- Formulate a model which represents a trained ReLU neural network as a P-split formulation in Julia
- Computational experiments: comparing the solution times of the big-M and P-split formulations

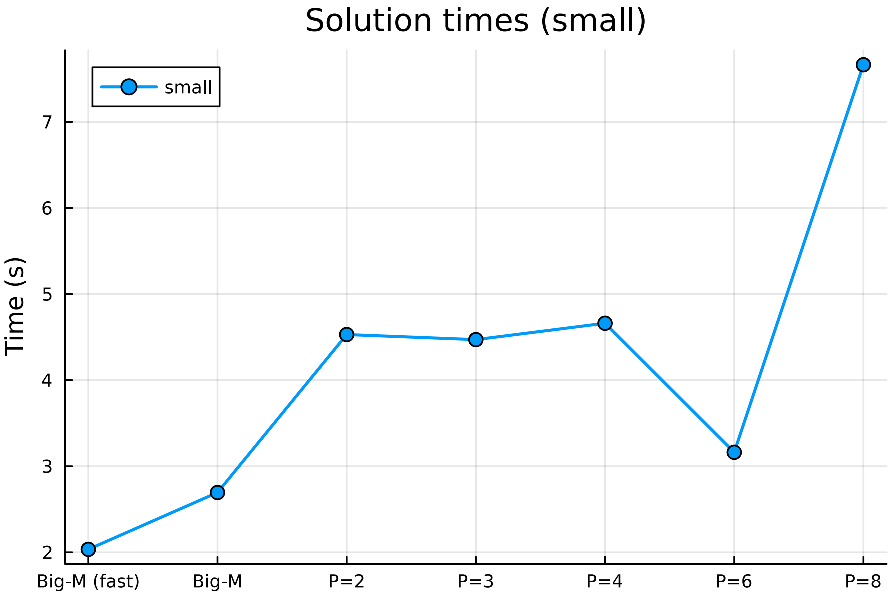
Computational tests

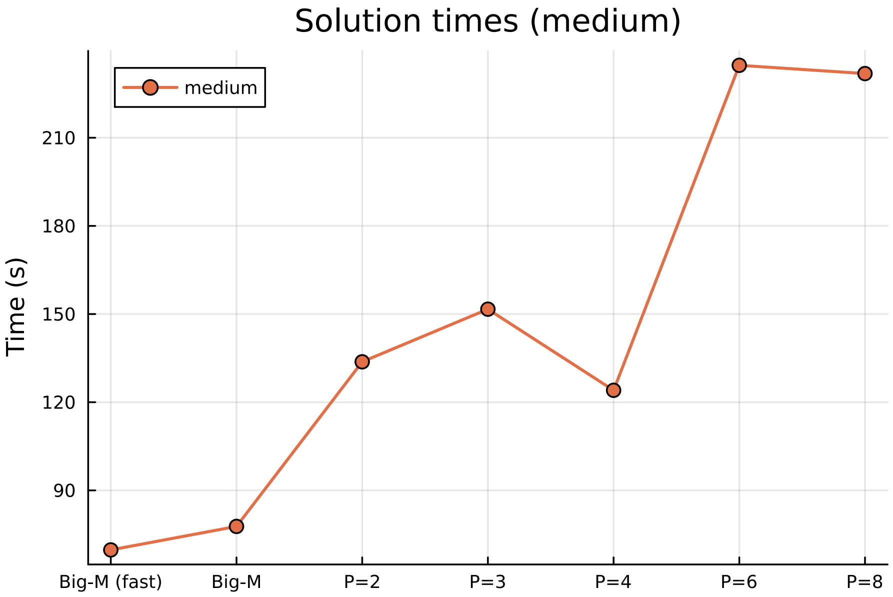
- Programming language: Julia
  - Neural networks: Flux.jl
  - 0-1 MILP formulation: Gogeta.jl
- Solver: Gurobi
- Data: Conrete Compressive Strength
- Maximizing the output of medium and large neural networks
- The tests were run 5 times for each NN and formulation and the average solution times and root relaxation objective values were documented

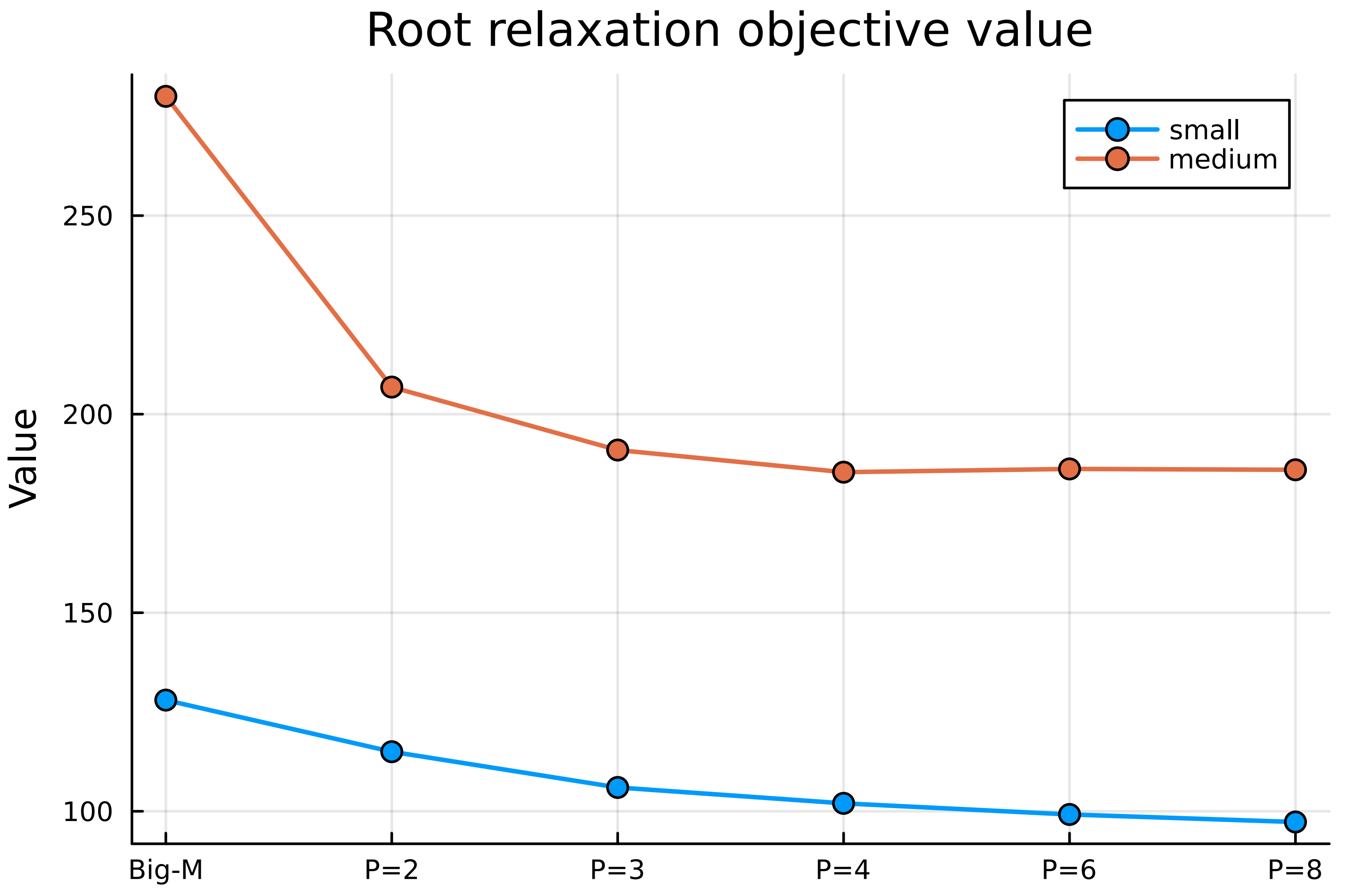
Network architectures

Size	Layers	Parameters	MAPE
medium	(8, 64, 32, 1)	2 689	11.94%
large	(8, 128, 64, 32, 1)	11 521	11.11%

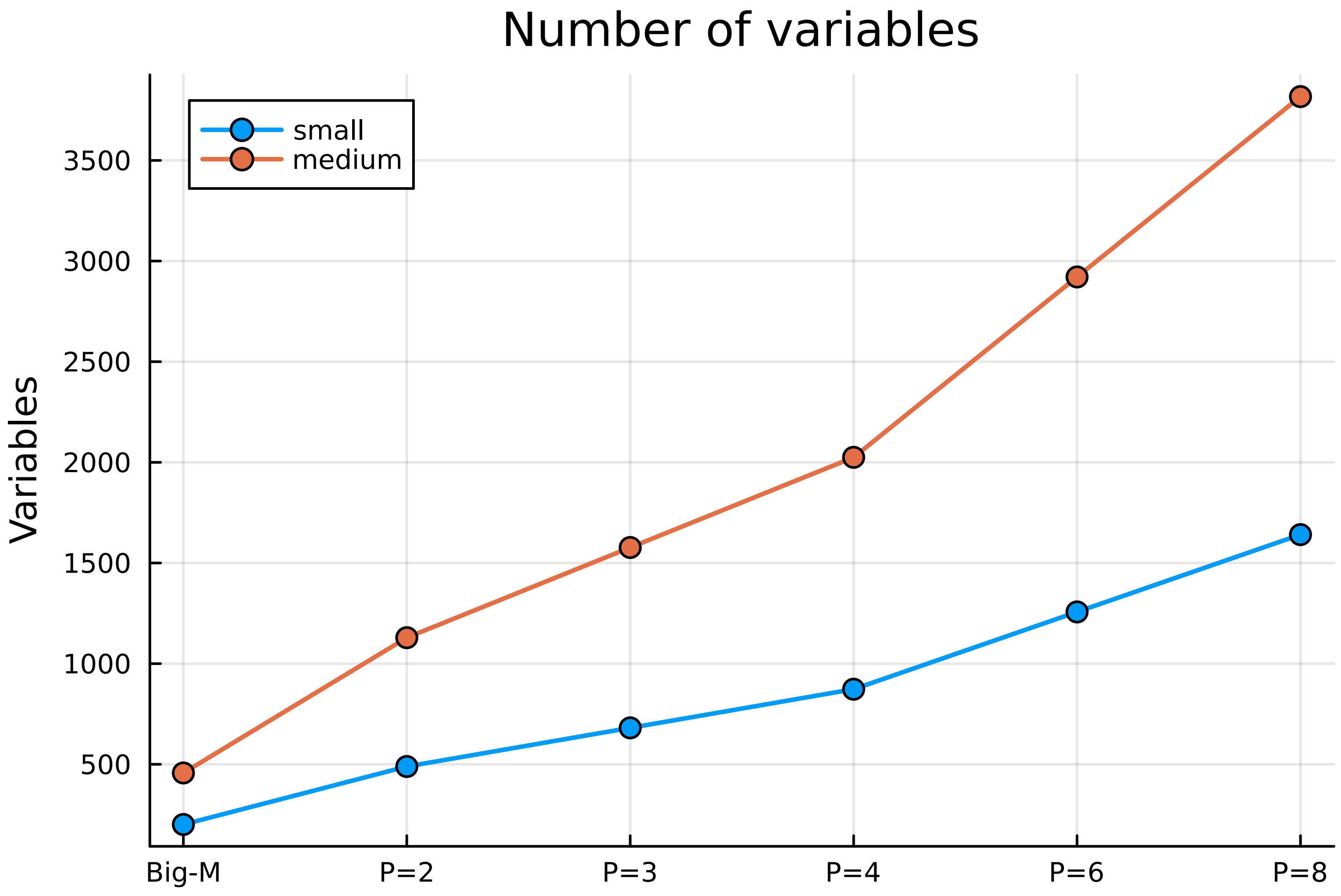
Solution times







Root relaxation objective value
Value
250
200
150
100
Big-M
P=2
P=3
P=4
P=6
P=8
small
medium


Number of variables
Variables
small
medium
3500
3000
2500
2000
1500
1000
500
Big-M
P=2
P=3
P=4
P=6
P=8

Results and conclusion

- The P-split formulation has no clear computational benefits over the big-M
- P-split provides a tighter linear relaxation than the big-M

- Limitations:
  - Only one dataset used
  - Effects of different network architectures should be explored
- Future work:
  - Find better performing formulation

References and literature

- Matteo Fischetti, Jason Jo, (2018). Deep neural networks and mixed integer linear optimization
- Jan Kronqvist, Ruth Misener, Calvin Tsay, (2022). P-split formulations: A class of intermediate formulations between big-M and convex hull for disjunctive constraints
- ReLU networks as surrogate models in mixed-integer linear programs (Grimstad and Andersson, 2019)
- Tsay, Calvin, et al. "Partition-based formulations for mixed-integer optimization of trained ReLU neural networks." *Advances in neural information processing systems* 34 (2021): 3068-3080.
- Gogeta.jl: https://gamma-opt.github.io/Gogeta.jl/dev/