Research Archive

NN and MIP

Jeff Linderoth, Changwon Lee

Motivated by *Rob Nowak, Fischetti and Jo*

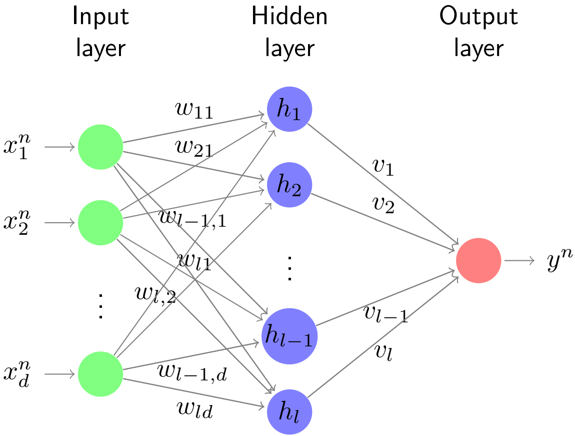
UW-Madison, August 2025

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Problem Setting

1. Given N datasets : $(\mathbf{x^1}, y^1)$,$(\mathbf{x^2}, y^2)$, $\cdots$ $(\mathbf{x^N}, y^N)$
2. Input : 'd'-dimensional
   $\mathbf{x^n} = (x_1^n, x_2^n, \ldots, x_d^n) \in \mathbb{R}^d$, $\quad \forall n \in [N]$
   ($\overline{\mathbf{x^n}} = (x_1^n, x_2^n, \ldots, x_d^n, 1) \in \mathbb{R}^{d+1}$)
3. w
4. $\mathbf{x^n}$ $\rightarrow$ Affine Combination : $\mathbf{w^k}\overline{\mathbf{x^n}}$ $\rightarrow$ Nonlinear Operator (ReLU) : $\rightarrow$ Activation of Neuron k

NN Formulation



1. $h_i^1 = max\{0, \mathbf{w} \cdot \mathbf{x^n} + b_i^1\} = max\{0, \mathbf{w} \cdot \overline{\mathbf{x^n}}\}$

MINLP Formulation 1

Decision vars:

$$\mathbf{W} \in \mathbb{R}^{l_1 \times d} : w_{ij} \in \mathbb{R}(i = 1, \ldots, l_1;\ j = 1, \ldots, d),$$
$$\mathbf{b} \in \mathbb{R}^{l_1} : b_i \in \mathbb{R}(i = 1, \ldots, l_1),$$
$$p_i^n, q_i^n \in \mathbb{R},$$
$$z_i^n \in \{0, 1\} \quad (n = 1, \ldots, N;\ i = 1, \ldots, l_1),$$
$$\mathbf{v} \in \mathbb{R}^{l_1} : v_i \in \mathbb{R}(i = 1, \ldots, l_1),$$
$$s_{ij} \in \mathbb{R} \quad (i = 1, \ldots, l_1;\ j = 1, \ldots, d),$$
$$t_{ij} \in \{0, 1\} \quad (i = 1, \ldots, l_1;\ j = 1, \ldots, d),$$

Objective:

$$\min \sum_{i=1}^{l_1} \sum_{j=1}^{d} t_{ij}$$

Constraints:

$$\sum_{j=1}^{d} w_{ij} x_j^n + b_i = p_i^n - q_i^n, \qquad \forall n \in [N], i \in [l_1]$$
$$p_i^n, q_i^n \geq 0, \qquad \forall n \in [N], i \in [l_1]$$
$$p_i^n \leq M(1 - z_i^n), \qquad \forall n \in [N], i \in [l_1]$$
$$q_i^n \leq M z_i^n, \qquad \forall n \in [N], i \in [l_1]$$
$$\sum_{i=1}^{l_1} p_i^n v_i = y^n, \qquad \forall n \in [N]$$
$$s_{ij} = w_{ij} v_i, \qquad \forall i \in [l_1], j \in [d]$$
$$-Mt_{ij} \leq s_{ij} \leq Mt_{ij}, \qquad \forall i \in [l_1], j \in [d]$$

Normalization of Single-Layer ReLU Networks

$Lemma 1. For any given NN weights/biases$ W, b,v,
there exists an (unique) alternate weights/biases $\mathbf{W'}, \mathbf{b'}, \mathbf{v'}$
such that $ReLU(\mathbf{W}\mathbf{x^n} + \mathbf{b})^T\mathbf{v} = ReLU(\mathbf{W'}\mathbf{x^n} + \mathbf{b'})^T\mathbf{v'} = y^n$ for all dataset $(\mathbf{x^n}, y)$, $n \in [N]$
and $v'_i \in \{-1, 0, +1\} \quad \forall i \in [l]$.

MILP Formulation by Lemma 1

Decision vars:

$$\mathbf{W} \in \mathbb{R}^{l \times d} : w_{ij} \in \mathbb{R} \quad (i = 1, \ldots, l;\ j = 1, \ldots, d),$$

$$\mathbf{b} \in \mathbb{R}^{l} : b_i \in \mathbb{R} \quad (i = 1, \ldots, l),$$

$$p_i^n, q_i^n \in \mathbb{R} \quad (n = 1, \ldots, N;\ i = 1, \ldots, l),$$

$$z_i^n \in \{0, 1\} \quad (n = 1, \ldots, N;\ i = 1, \ldots, l),$$

$$\mathbf{v} \in \{-1, 0, +1\}^{l} : v_i^{-1}, v_i^{0}, v_i^{+1} \in \{0, 1\} \quad (i = 1, \ldots, l),$$

$$s_{ij} \in \mathbb{R} \quad (i = 1, \ldots, l_1;\ j = 1, \ldots, d),$$

$$t_{ij} \in \{0, 1\} \quad (i = 1, \ldots, l_1;\ j = 1, \ldots, d),$$

Objective:

$$\min \sum_{i=1}^{l_1} \sum_{j=1}^{d} t_{ij}$$

Constraints:

$$\sum_{j=1}^{d} w_{ij} x_j^n + b_i = p_i^n - q_i^n, \qquad \forall n \in [N], i \in [l_1]$$

$$p_i^n, q_i^n \geq 0, \qquad \forall n \in [N], i \in [l_1]$$

$$p_i^n \leq M(1 - z_i^n), \qquad \forall n \in [N], i \in [l_1]$$

$$q_i^n \leq M z_i^n, \qquad \forall n \in [N], i \in [l_1]$$

$$\sum_{i=1}^{l} p_i^n (-v_i^{-1} + v_i^{+1}) = y^n, \qquad \forall n \in [N]$$

$$s_{ij} = w_{ij} v_i, \qquad \forall i \in [l_1], j \in [d]$$

$$-Mt_{ij} \leq s_{ij} \leq Mt_{ij}, \qquad \forall i \in [l_1], j \in [d]$$

References I