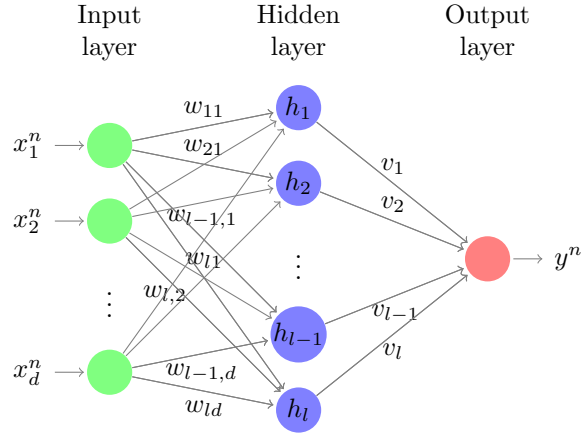


Normalization of Single-Layer ReLU Networks

Consider a single-hidden layer ReLU Neural Network: $ReLU(\mathbf{W}\mathbf{x} + \mathbf{b})^T \mathbf{v} = y$



Where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \in \mathbb{R}^d, \quad y \in \mathbb{R}$$

and

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1d} \\ w_{21} & w_{22} & \cdots & w_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ w_{l1} & w_{l2} & \cdots & w_{ld} \end{bmatrix} \in \mathbb{R}^{l \times d}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_l \end{bmatrix} \in \mathbb{R}^l, \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_l \end{bmatrix} \in \mathbb{R}^l.$$

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

Proof. For a given \mathbf{v} , let $I \subseteq [l]$ be the index set where $v_i = 0 \quad \forall i \in I$ and $v_i \neq 0 \quad \forall i \in [l]/I$.

Let \mathbf{v}' be $v'_i = 0 \quad \forall i \in I$ and $v'_i = \frac{v_i}{|v_i|} \in \{-1, +1\} \quad \forall i \in [l]/I$.

Let \mathbf{W}' be $w'_{ij} = w_{ij}$ (or any arbitrary scalar) $\quad \forall i \in I, \forall j \in [d]$ and $w'_{ij} = |v_i|w_{ij} \quad \forall i \in [l]/I, \forall j \in [d]$.

Let \mathbf{b}' be $b'_i = b_i$ (or any arbitrary scalar) $\quad \forall i \in I$ and $b'_i = |v_i|b_i \quad \forall i \in [l]/I$.

Then for an arbitrary dataset (\mathbf{x}^n, y^n) ,

Since $v_i = v'_i = 0 \quad \forall i \in I$, then $ReLU(\mathbf{W}\mathbf{x}^n + \mathbf{b})_i v_i = ReLU(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')_i v'_i = 0 \quad \forall i \in I$.

Since $w'_{ij} = |v_i|w_{ij} \quad \forall i \in [l]/I, \forall j \in [d]$ and $b'_i = |v_i|b_i \quad \forall i \in [l]/I$,

then $(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')_i = \mathbf{W}'^T_i \mathbf{x}^n + \mathbf{b}'_i = |v_i|\mathbf{W}^T_i \mathbf{x}^n + |v_i|\mathbf{b}_i = |v_i|(\mathbf{W}\mathbf{x}^n + \mathbf{b})_i \quad \forall i \in [l]/I$.

Since $(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')_i = |v_i|(\mathbf{W}\mathbf{x}^n + \mathbf{b})_i \quad \forall i \in [l]/I$,

if $(\mathbf{W}\mathbf{x}^n + \mathbf{b})_i > 0$ then $(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')_i > 0$ ($\because v_i \neq 0 \quad \forall i \in [l]/I$) and $ReLU(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')_i = (\mathbf{W}'\mathbf{x}^n + \mathbf{b}')_i = |v_i|(\mathbf{W}\mathbf{x}^n + \mathbf{b})_i = |v_i|ReLU(\mathbf{W}\mathbf{x}^n + \mathbf{b})_i$.

If $(\mathbf{W}\mathbf{x}^n + \mathbf{b})_i \leq 0$ then $(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')_i \leq 0$ ($\because v_i \neq 0 \quad \forall i \in [l]/I$) and $ReLU(\mathbf{W}\mathbf{x}^n + \mathbf{b})_i = ReLU(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')_i = 0$.

Therefore $ReLU(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')_i = |v_i|ReLU(\mathbf{W}\mathbf{x}^n + \mathbf{b})_i \quad \forall i \in [l]/I$

and $ReLU(\mathbf{W}\mathbf{x}^n + \mathbf{b})_i v_i = \frac{ReLU(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')_i}{|v_i|} v_i = ReLU(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')_i v'_i = 0 \quad \forall i \in [l]/I$.

Thus, since $ReLU(\mathbf{W}\mathbf{x}^n + \mathbf{b})_i v_i = ReLU(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')_i v'_i = 0 \quad \forall i \in [l]$,

then $ReLU(\mathbf{W}\mathbf{x}^n + \mathbf{b})^T \mathbf{v} = ReLU(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')^T \mathbf{v}' = y^n$

□

Corollary 1. For given dataset (\mathbf{x}^n, y^n) , $n \in [N]$,

if the problem of finding $\mathbf{W}, \mathbf{b}, \mathbf{v}$ that satisfies $ReLU(\mathbf{W}\mathbf{x}^n + \mathbf{b})^T \mathbf{v} = y^n$ for all dataset (\mathbf{x}^n, y^n) , $n \in [N]$, is feasible, then finding $\mathbf{W}', \mathbf{b}', \mathbf{v}'$ such that $ReLU(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')^T \mathbf{v}' = y^n$ for all dataset (\mathbf{x}^n, y) , $n \in [N]$ and $v'_i \in \{-1, 0, +1\} \quad \forall i \in [l]$ is also feasible.

Corollary 2. For given dataset (\mathbf{x}^n, y^n) , $n \in [N]$,

if there doesn't exist any $\mathbf{W}', \mathbf{b}', \mathbf{v}'$ such that $ReLU(\mathbf{W}'\mathbf{x}^n + \mathbf{b}')^T \mathbf{v}' = y^n$ for all dataset (\mathbf{x}^n, y) , $n \in [N]$ and $v'_i \in \{-1, 0, +1\} \quad \forall i \in [l]$, then also there doesn't exist any $\mathbf{W}, \mathbf{b}, \mathbf{v}$ that satisfies $ReLU(\mathbf{W}\mathbf{x}^n + \mathbf{b})^T \mathbf{v} = y^n$ for all dataset (\mathbf{x}^n, y^n) , $n \in [N]$.