Normalization of Single-Layer ReLU Networks

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

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 given NN weights/biases* $\mathbf{W},\mathbf{b},\mathbf{v}$*,*
there exists an (unique) alternate weights/biases $\mathbf{W}',\mathbf{b}',\mathbf{v}'$
such that $ReLU(\mathbf{Wx^n}+\mathbf{b})^T\mathbf{v}=ReLU(\mathbf{W'x^n}+\mathbf{b}')^T\mathbf{v}'=y^n$ *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 $w_{ij}'=w_{ij}\quad\forall i\in I,\forall j\in[d]$ and $w_{ij}'=\frac{w_{ij}}{|v_i|}\quad\forall i\in[l]/I,\forall j\in[d]$.

Let $\mathbf{b}'$ be $b_i'=b_i\quad\forall i\in I$ and $b_i'=\frac{b_i}{|v_i|}\quad\forall i\in[l]/I$.

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

$(\mathbf{Wx^n}+\mathbf{b})_i=(\mathbf{W'x^n}+\mathbf{b}')_i\quad\forall i\in I$ and therefore $ReLU(\mathbf{Wx^n}+\mathbf{b})_i=ReLU(\mathbf{W'x^n}+\mathbf{b}')_i\quad\forall i\in I$.
Then $ReLU(\mathbf{Wx^n}+\mathbf{b})_iv_i=ReLU(\mathbf{W'x^n}+\mathbf{b}')_iv_i'=0\quad\forall i\in I$.

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

$\square$