DDA4250 Assignment 1

CHEN Ang (118010009)

Notation

For conciseness we introduce the following notation.

- $\mathbf{R}$: multidimensional rectifier functions of suitable dimensions
- $[a..b] := \{x \in \mathbb{Z} : a \leq x \leq b\}$

2.1.2

With $\boldsymbol{\theta} = (1, -1, 0, 0, 1, 1, 0) \in \mathbb{R}^7, l_1 = l_L = 2$,

$$\mathcal{N}^{\boldsymbol{\theta},1}_{\mathbf{R},\mathrm{id}_{\mathbb{R}}}(x) = \mathbf{R}(x) + \mathbf{R}(-x) = \begin{cases} \mathbf{R}(x) = x, & \text{if } x \geq 0 \\ \mathbf{R}(-x) = -x, & \text{otherwise} \end{cases} = |x|$$

$$\mathbf{d} = 7 = 2l_1 + \left[\sum_{k=2}^{L} l_k \left(l_{k-1} + 1\right)\right] + l_L + 1$$

2.1.3

Definition. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be piecewise linear with a finite (interval) partition if there exist $n \in \mathbb{N}$ and an interval partition $P = \{p_1, \ldots, p_n\}$ of $\mathbb{R}$ such that $f(x) = f_i(x)$ is an affine transform on each of the interval $p_i \in P$. It is required that $p_i \neq \emptyset$ for all $p_i \in P$.

Lemma. Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be two piecewise linear functions with finite partitions. Then the function $\mathcal{L} : \mathbb{R} \to \mathbb{R}$ given by $x \to c_1 f(x) + c_2 g(x) + d$ is also piecewise linear with a finite partition.

Proof of Lemma. Let $P = \{p_1, \ldots, p_m\}$ be a partition of $\mathbb{R}$ such that in each interval $p_i \in P$, $f(x) = f_i(x)$ is an affine transform. Similarly let $Q = \{q_1, \ldots, q_n\}$ be a partition of $\mathbb{R}$ such that in each interval $q_i \in Q$, $g(x) = g_j(x)$ is affine. Clearly,

$$S = \{p \cap q : p \in P, q \in Q\} \setminus \emptyset$$

is also a (finite) partition of $\mathbb{R}$. Moreover, for each interval $s \in S$, there exist $(i, j) \in [1..m] \times [1..n]$ such that $s = p_i \cap q_j$, and thus for all $x \in s$,

$$\mathcal{L}(x) = c_1 f(x) + c_2 g(x) + d = c_1 f_i(x) + c_2 g_i(x) + d$$

is affine. Hence $\mathcal{L}$ is piecewise linear with a finite partition.

There is no such $\mathcal{N}$. To see that, let $\mathcal{N}^{\boldsymbol{\theta},1}_{\mathbf{R},\ldots\mathbf{R},\mathrm{id}_{\mathbb{R}}}$ be given. We can show by induction that for any layer $k \in [0..L+1]$ of the network, the neurons of the layer $\boldsymbol{x}^{(k)}$ are all piecewise linear functions of x with finite partitions.

Layer 0 is just the input neuron with value x, which is linear on the entire $\mathbb{R}$.

Suppose that all neurons $\boldsymbol{x}^{(k)}$ of some layer $k \in [0..L]$ are piecewise linear with finite partitions. Consider neurons on the next layer:

$$\boldsymbol{x}^{(k+1)} = \mathbf{R} \circ \mathcal{A}^{(k)} \left(\boldsymbol{x}^{(k)}\right) = \mathbf{R}\left(\boldsymbol{W}^{(k)}\boldsymbol{x}^{(k)} + \boldsymbol{b}^{(k)}\right)$$

Equivalently for all $i \in [1..l_k]$,

$$\boldsymbol{x}_i^{(k+1)} = \mathbf{R}\left(\boldsymbol{w}_i^\top \boldsymbol{x}^{(k)} + \boldsymbol{b}_i^{(k)}\right)$$

where $\boldsymbol{w}_i^\top$ denotes the i-th row of $\boldsymbol{W}^{(k)}$. From the assumption that $\boldsymbol{x}^{(k)}$ are all piecewise linear with finite partitions, we know by Lemma that $\boldsymbol{w}_i^\top \boldsymbol{x}^{(k)} + \boldsymbol{b}_i^{(k)}$ is also piecewise linear with some finite partition $P = \{p_1, \ldots, p_n\}$. Thus on each interval $p_j \in P$, $f_j(x) := \boldsymbol{w}_i^\top \boldsymbol{x}^{(k)} + \boldsymbol{b}_i^{(k)}$ is affine. Define for each $j \in [1..n]$ the set $\Omega_j^+ := \{x : x \in p_j, f_j(x) \geq 0\}$ and $\Omega_j^- := p_j \backslash \Omega_j^+$. It holds that for $x \in p_j$, $j \in [1..n]$,

$$\mathbf{R} \circ f_j(x) = \begin{cases} f_j(x), & \text{if } x \in \Omega_j^+ \\ 0, & \text{if } x \in \Omega_j^- \end{cases}$$

is affine in either set of the partition $\{\Omega_j^+, \Omega_j^-\}$. Further, since $f_j(x)$ is affine, both Ω_j^+ and Ω_j^- must be intervals (possibly empty). Hence

$$Q := \left\{\Omega_j^+ : j \in [1..n]\right\} \cup \left\{\Omega_j^- : j \in [1..n]\right\} \setminus \emptyset$$

forms a finite interval partition of $\mathbb{R}$ such that $\boldsymbol{x}_i^{(k+1)}$ is affine on any interval $q \in Q$. In other words, $\boldsymbol{x}_i^{(k+1)}$ is piecewise linear with a finite partition for all i, completing the induction.

Applying this result on layer $L+1$ shows that the output $\mathcal{N}^{\boldsymbol{\theta},1}_{\mathbf{R},\ldots\mathbf{R},\mathrm{id}_{\mathbb{R}}}(x) = \boldsymbol{x}_{L+1}$ must be piecewise linear in x with a finite partition, which cannot be e^x.

2.1.4

With $\boldsymbol{\theta} = (1, -1, 0, 1, 0, -1, 0, 0, 0, 1, 1, -1, 0) \in \mathbb{R}^{13}, l_1 = l_L = 3$, for all $(x, y) \in \mathbb{R}^2$

$$\begin{aligned} \mathcal{N}^{\boldsymbol{\theta},2}_{\mathbf{R},\mathrm{id}_{\mathbb{R}}}(x, y) &= \mathbf{R}(x - y) + \mathbf{R}(y) - \mathbf{R}(-y) \\ &= \mathbf{R}(x - y) + y \\ &= \begin{cases} x - y + y = x, & \text{if } x \geq y \\ 0 + y = y, & \text{otherwise} \end{cases} \\ &= \max\{x, y\} \end{aligned}$$

$$\mathbf{d} = 13 = 3l_1 + \left[\sum_{k=2}^{L} l_k\left(l_{k-1} + 1\right)\right] + l_L + 1$$

2.1.4

With $\boldsymbol{\theta} = (1, -1, 0, 0, 1, 0, 0, 1, 0, 0, 1, -1, 0) \in \mathbb{R}^{13}, l_1 = 2, l_2 = 2$, for all $x \in \mathbb{R}$

$$\begin{aligned} \mathcal{N}^{\boldsymbol{\theta},1}_{\mathbf{R},\mathbf{R},\mathrm{id}_{\mathbb{R}}}(x) &= \mathbf{R} \circ \mathbf{R}(x) - \mathbf{R} \circ \mathbf{R}(-x) \\ &= \mathbf{R}(x) - \mathbf{R}(-x) \\ &= x \end{aligned}$$

$$\mathbf{d} = 13 = 2l_1 + l_1 l_2 + 2l_2 + 1$$

where the second equality comes from the fact that $\mathbf{R}(x) \geq 0$.