Odvození (7)-(12) + analýza

Stanislav Novotný

December 21, 2020

Mějme komplexní matici vah $\mathbf{W}$, kterou můžeme napsat jako součet její reálné a imaginarní části $\mathbf{W} = \mathbf{W}^r + i\mathbf{W}^i$. Dále uvažujme logaritmus definovaný na komplexních číslech $\log x = \log r + i\Theta = \log r + i\pi k$, kde $r = |x|$. Vyjdeme z rovnice (7) z [1]

$$
\begin{aligned}
\mathbf{z} &= \exp(\mathbf{W}\log\mathbf{x}) = \exp((\mathbf{W}^r + i\mathbf{W}^i)(\log\mathbf{r} + i\pi\mathbf{k})) \\
&= \exp(\mathbf{W}^r\log\mathbf{r})\exp(i\pi\mathbf{W}^r\mathbf{k})\exp(i\mathbf{W}^i\log\mathbf{r})\exp(-\pi\mathbf{W}^i\mathbf{k}) \\
&= \exp(\mathbf{W}^r\log\mathbf{r})(\cos(\pi\mathbf{W}^r\mathbf{k}) + i\sin(\pi\mathbf{W}^r\mathbf{k}))(\cos(\mathbf{W}^i\log\mathbf{r}) + i\sin(\mathbf{W}^i\log\mathbf{r}))\exp(-\pi\mathbf{W}^i\mathbf{k}) \\
&= \exp(\mathbf{W}^r\log\mathbf{r})\exp(-\pi\mathbf{W}^i\mathbf{k})\Bigg[\frac{1}{2}\cos(\pi\mathbf{W}^r\mathbf{k} + \mathbf{W}^i\log\mathbf{r}) + \frac{1}{2}\cos(\pi\mathbf{W}^r\mathbf{k} - \mathbf{W}^i\log\mathbf{r}) + i\Bigg(\cos(\pi\mathbf{W}^r\mathbf{k})\sin(\mathbf{W}^i\log\mathbf{r}) \\
&+ \sin(\pi\mathbf{W}^r\mathbf{k})\cos(\mathbf{W}^i\log\mathbf{r})\Bigg) - \frac{1}{2}\cos(\pi\mathbf{W}^r\mathbf{k} - \mathbf{W}^i\log\mathbf{r}) + \frac{1}{2}\cos(\pi\mathbf{W}^r\mathbf{k} + \mathbf{W}^i\log\mathbf{r})\Bigg] \\
&= \exp(\mathbf{W}^r\log\mathbf{r})\exp(-\pi\mathbf{W}^i\mathbf{k})\Bigg(\cos(\pi\mathbf{W}^r\mathbf{k} + \mathbf{W}^i\log\mathbf{r}) + i\Bigg(\cos(\pi\mathbf{W}^r\mathbf{k})\sin(\mathbf{W}^i\log\mathbf{r}) + \sin(\pi\mathbf{W}^r\mathbf{k})\cos(\mathbf{W}^i\log\mathbf{r})\Bigg)\Bigg) \\
&= \exp(\mathbf{W}^r\log\mathbf{r})\exp(-\pi\mathbf{W}^i\mathbf{k})\cos(\pi\mathbf{W}^r\mathbf{k} + \mathbf{W}^i\log\mathbf{r}) + i\Bigg(\cos(\pi\mathbf{W}^r\mathbf{k})\sin(\mathbf{W}^i\log\mathbf{r}) \\
&+ \sin(\pi\mathbf{W}^r\mathbf{k})\cos(\mathbf{W}^i\log\mathbf{r})\Bigg)\exp(\mathbf{W}^r\log\mathbf{r})\exp(-\pi\mathbf{W}^i\mathbf{k})
\end{aligned}
$$

Budeme uvažovat pouze reálnou část a rozepíšeme si vrstvu NPU pro n vstupů a 1 výstup Nechť tedy

$$
\mathbf{x} \in \mathbb{R}^{n\times 1},\ \mathbf{W} \in \mathbb{R}^{1\times n},\ \mathbf{r} = |\mathbf{x}|,\ k_i = \begin{cases} 1, & x_i \leq 0 \\ 0, & x_i > 0 \end{cases}
$$

$$
\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \mathbf{W} = \mathbf{W}^r + i\mathbf{W}^i = (w_1^r \cdots w_n^r) + i\left(w_1^i \cdots w_n^i\right)
$$

$$
\begin{aligned}
\mathbf{z} &= \exp(\mathbf{W}^r \log \mathbf{r}) \exp(-\pi \mathbf{W}^i \mathbf{k}) \cos(\pi \mathbf{W}^r \mathbf{k} + \mathbf{W}^i \log \mathbf{r}) \\
&= \exp\left(\sum_{s=1}^{n} w_s^r \log r_s\right) \exp\left(-\pi \sum_{s=1}^{n} w_s^i k_s\right) \cos\left(\pi \sum_{s=1}^{n} w_s^r k_s + \sum_{s=1}^{n} w_s^i \log r_s\right) \\
&= \exp(w_1^r \log r_1) \dots \exp(w_n^r \log r_n) \exp\left(-\pi \sum_{s=1}^{n} w_s^i k_s\right) \cos\left(\pi \sum_{s=1}^{n} w_s^r k_s + \sum_{s=1}^{n} w_s^i \log r_s\right) \\
&= r_1^{w_1^r} \dots r_n^{w_n^r} \exp\left(-\pi \sum_{s=1}^{n} w_s^i k_s\right) \cos\left(\pi \sum_{s=1}^{n} w_s^r k_s + \sum_{s=1}^{n} w_s^i \log r_s\right)
\end{aligned}
$$

Pokud má vektor $\mathbf{x}$ všechny složky nekladné tj. $x_i \leq 0 \ \forall i \in 1 \dots n$, výstup bude mít tvar

$$
\mathbf{z} = r_1^{w_1^r} \dots r_n^{w_n^r} \exp(-\pi \mathbb{W}^i) \cos\left(\pi \mathbb{W}^r + \sum_{s=1}^{n} \log r_s^{w_s^i}\right)
$$

Naopak, pokud $x_i > 0 \ \forall i \in 1 \dots n$, získáme

$$
\mathbf{z} = r_1^{w_1^r} \dots r_n^{w_n^r} \cos\left(\sum_{s=1}^{n} \log r_s^{w_s^i}\right)
$$

Věta 0.1. *Kombinací NaiveNPU a Dense vrstvy jsme schopni vyjádřit jakýkoliv polynom* $p(x) = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n \sim$ *Chain(NaiveNPU(1,n), Dense(n,1,identity)).*

Důkaz. Nyní uvažujeme 1 vstup a n výstupů vrstvy NaiveNPU, mocniny polynomu jsou přirozená čísla.

$$
x \in \mathbb{R}^{1\times 1}, \ \mathbf{W} \in \mathbb{N}^{n\times 1}, \ r = |x|
$$

Důkaz rozdělíme na dvě části podle hodnoty vstupu.

Pro kladný vstup x bude mít výstup NPU tvar

$$
\exp(\mathbf{W}^r \log \mathbf{r}) \cos(\mathbf{W}^i \log \mathbf{r})
$$

$\mathbf{W}$ by tedy po iteraci měla vypadat takto

$$
\mathbf{W} = \mathbf{W}^r + i\mathbf{W}^i = \begin{pmatrix} 1 \\ \vdots \\ n \end{pmatrix} + i \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}
$$

a výstup NaiveNPU takto

$$
\mathbf{z} = \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}
$$

Matice vah $\mathbf{A} \in \mathbb{R}^{1\times n}$ vrstvy Dense pro výstup NaiveNPU pak udělá lineární kombinaci a přidá bias, neboli v tomto případě absolutní člen polynomu.

$$
\mathbf{A} = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}, \ b = a_0
$$

Pro nekladný vstup $-x$ nabude výstup NaiveNPU tvaru

$$
\exp(\mathbf{W}^r \log \mathbf{r}) \exp(-\pi \mathbf{W}^i) \cos(\pi \mathbf{W}^r + \mathbf{W}^i \log \mathbf{r})
$$

Matice $\mathbf{W}$ bude mít stejný tvar jako v předešlém případě a výstup tedy bude vypadat následovně

$$\mathbf{z} = \begin{pmatrix} x^1 \\ -x^2 \\ x^3 \\ \vdots \\ \pm x^n \end{pmatrix}$$

v závislosti na lichosti, respektive sudosti n. Matice Dense vrstvy se nyní pozmění na tvar

$$\mathbf{A} = \begin{pmatrix} a_1 & -a_2 & a_3 & \cdots & \pm a_n \end{pmatrix},\ b = a_0$$

ve stejné závislosti. $\square$

Příklad 1. Polynom $2x^3+5x^2+5$ bychom dostali pomocí Chain(NaiveNPU(1, 2), Dense(2, 1, identity)) následovně $x \xrightarrow{NaiveNPU(1,2)} x^3, x^2 \xrightarrow{Dense(2,1,id)} 2x^3 + 5x^2 + 5$.

Kombinací více NaiveNPU a Dense vrstev jsme schopni vytvářet složitější zobrazení než polynomy, např. racionální lomené funkce.

Věta 0.2. *Mějme $H(x) = \frac{p(x)}{q(x)}$, kde $p(x) = \sum_{i=0}^{n} a_i x^i$, $q(x) = \sum_{i=0}^{m} b_i x^i$ nenulový. Pak $H(x)$ vznikne jako Chain(NaiveNPU(1, n+m), Dense(n+m, 2, identity), NaiveNPU(2, 1))*

Důkaz. Nechť

$$x \in \mathbb{R}^{1\times 1},\ \mathbf{W}_1 \in \mathbb{Z}^{n+m\times 1},\ \mathbf{A} \in \mathbb{R}^{2\times n+m},\ \mathbf{W}_2 \in \mathbb{Z}^{1\times 2},\ r = |x|$$

postup bude obdobný jako v důkazů (0.1). Nejprve rozebereme případ kladného vstupu x. Výstup první NaiveNPU vrstvy bude mít tvar

$$\mathbf{z}_1 = \exp(\mathbf{W}^r \log \mathbf{r}) \cos(\mathbf{W}^i \log \mathbf{r})$$

Iterací získá $\mathbf{W}_1$ tvar

$$\mathbf{W}_1 = \mathbf{W}_1^r + i\mathbf{W}_1^i = \begin{pmatrix} 1 \\ \vdots \\ n \\ 1 \\ \vdots \\ m \end{pmatrix} + i \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Výstup tedy bude mít tvar

$$\mathbf{z}_1 = \begin{pmatrix} x^1 \\ \vdots \\ x^n \\ x^1 \\ \vdots \\ x^m \end{pmatrix}$$

Pro kýžený výsledek musí matice vah vrstvy Dense nabýt tvaru

$$\mathbf{A} = \begin{pmatrix} a_1 & \cdots & a_n & 0 & \cdots & 0 \\ 0 & \cdots & 0 & b_1 & \cdots & b_m \end{pmatrix},\ b = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix}$$

Působením Dense vrstvy na výstup první NaiveNPU vrstvy dostaneme

$$\mathbf{z}_2 = \begin{pmatrix} p(x) \\ q(x) \end{pmatrix}$$

Matice vah v poslední vrstvě bude mít tvar

$$\mathbf{W}_2 = \mathbf{W}_2^r + i\mathbf{W}_2^i = \begin{pmatrix} 1 & -1 \end{pmatrix} + i \begin{pmatrix} 0 & 0 \end{pmatrix}$$

Takto již získáme $H(x)$
tj. $x \xrightarrow{NPU(1,n+m)} x_1, \ldots, x_{n+m} \xrightarrow{Dense(n+m,2,id)} f(x), g(x) \xrightarrow{NPU(2,1)} H(x)$.
Pro nekladný vstup by byl postup dokazování obdobný $\square$

References

[1] Niklas Heim, Tomáš Pevný, and Václav Šmídl. Neural power units. *arXiv preprint arXiv:2006.01681*, 2020.