Odvození (7)-(12) + analýza

Stanislav Novotný

December 10, 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

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

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

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

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

Důkaz. Ze vstupních dat $\mathbf{x}$ vrstva NPU udělá n výstupů, z nichž každý může být umocněn na libovolnou mocninu, tyto data přijme vrstva Dense, která každý tento vstup přeškáluje a přidá mu určitý posun tj. Dense vstup zpracuje jako $I.(m.W * x + m.b)$, kde I představuje identickou aktivační funkci, m je koeficient, W matice vah a b je bias. □

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

Kombinací více NPU a Dense vrstev jsme schopni vytvářet složitější zobrazení než polynomy, např. racionální lomené funkce. 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(NPU(1, n+m), Dense(n+m, 2, identity), NPU(2, 1)) tj. $x \xrightarrow{NPU(1,n+m)} x_1, \dots, x_{n+m} \xrightarrow{Dense(n+m,2,id)} f(x), g(x) \xrightarrow{NPU(2,1)} H(x)$

References

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