Product of As Expanded

Theorem

$$\overset{[N]}{\bar{A}_\ell}\,\overset{[N]}{\bar{A}_{\ell-1}}\dots\overset{[N]}{\bar{A}_3}\,\overset{[N]}{\bar{A}_2}$$

$$= \mathrm{pow}\Big(\overbrace{\Big[\sum_{S\subseteq\{\ell,\ell-1,\dots,3,2\}} \exp\big(|S|\overset{[1]}{B_\Delta[e]}\big)\exp\big(\overset{[E]}{W_\Delta[:,e]}\cdot\sum_{i\in S}\overset{[E]}{x_i}\big)\Big]}^{[1]}, \overset{[N]}{A[e]}\Big)$$

Context

We have

$$\overset{[1]}{\Delta_{\ell,e}} = \mathrm{softplus}(\overset{[E]}{x_\ell}\cdot \overset{[E]}{W_\Delta[:,e]} + \overset{[1]}{B_\Delta[e]})$$

$$\overset{[N]}{\bar{A}_{\ell,e}} = \exp(\overset{[1]}{\Delta_{\ell,e}}\overset{[N]}{A[e]})$$

$$\overset{[N]}{B_\ell} = \overset{[N,E]}{W_B}\overset{[E]}{x_\ell}$$

$$\overset{[N]}{\bar{B}_{\ell,e}} = \overset{[1]}{\Delta_{\ell,e}}\overset{[N]}{B_\ell}$$

$$\overset{[N]}{h_{\ell,e}} = \overset{[N]}{\bar{A}_{\ell,e}}\overset{[N]}{h_{\ell-1,e}} + \overset{[N]}{\bar{B}_{\ell,e}}\overset{[1]}{x_{\ell,e}}$$

$$\overset{[N]}{C_\ell} = \overset{[N,E]}{W_C}\overset{[N]}{x_\ell}$$

$$\overset{[1]}{y_{\ell,e}} = \overset{[N]}{h_{\ell,e}}\cdot\overset{[N]}{C_\ell}$$

$$\overset{[N]}{h_{\ell,e}} = \exp(\overset{[1]}{\Delta_{\ell,e}}\overset{[N]}{A[e]})\overset{[N]}{h_{\ell-1,e}} + \overset{[1]}{\Delta_{\ell,e}}\overset{[N,E]}{W_B}\overset{[E]}{x_\ell}\overset{[1]}{x_{\ell,e}}$$

Lets expand the recurrance out a bit(temporairly omitting the e and b index for readability)

$$h_1 = \overset{[N]}{\bar{A}_1}\overset{[N]}{0} + \overset{[N]}{\bar{B}_1}\overset{[1]}{x_{1,e}} = \overset{[N]}{\bar{B}_1}\overset{[1]}{x_{1,e}}$$

$$h_2 = \overset{[N]}{\bar{A}_2}\overset{[N]}{h_1} + \overset{[N]}{\bar{B}_1}\overset{[1]}{x_{1,e}} = \overset{[N]}{\bar{A}_2}\overset{[N]}{\bar{B}_1}\overset{[1]}{x_{1,e}} + \overset{[N]}{\bar{B}_2}\overset{[1]}{x_{2,e}}$$

$$h_3 = \overset{[N]}{\bar{A}_3}\overset{[N]}{h_2} + \overset{[N]}{\bar{B}_3}\overset{[1]}{x_{3,e}} = \overset{[N]}{\bar{A}_3}\overset{[N]}{\bar{A}_2}\overset{[N]}{\bar{B}_1}\overset{[1]}{x_{1,e}} + \overset{[N]}{\bar{A}_3}\overset{[N]}{\bar{B}_2}\overset{[1]}{x_{2,e}} + \overset{[N]}{\bar{B}_3}\overset{[1]}{x_{3,e}}$$

$$h_4 = \overset{[N]}{\bar{A}_4}\overset{[N]}{h_3} + \overset{[N]}{\bar{B}_4}\overset{[1]}{x_{4,e}} = \overset{[N]}{\bar{A}_4}\overset{[N]}{\bar{A}_3}\overset{[N]}{\bar{A}_2}\overset{[N]}{\bar{B}_1}\overset{[1]}{x_{1,e}} + \overset{[N]}{\bar{A}_4}\overset{[N]}{\bar{A}_3}\overset{[N]}{\bar{B}_2}\overset{[1]}{x_{2,e}} + \overset{[N]}{\bar{A}_4}\overset{[N]}{\bar{B}_3}\overset{[1]}{x_{3,e}} + \overset{[N]}{\bar{B}_4}\overset{[1]}{x_{4,e}}$$

$$\ldots$$

$$h_\ell =$$

$$\overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \ldots \overset{[N]}{\bar{A}}_3 \overset{[N]}{\bar{A}}_2 \overset{[N]}{\bar{B}}_1 \overset{[1]}{x}_{1,e} +$$

$$\overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \ldots \overset{[N]}{\bar{A}}_4 \overset{[N]}{\bar{A}}_3 \overset{[N]}{\bar{B}}_2 \overset{[1]}{x}_{2,e} +$$

$$\vdots$$

$$\overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \overset{[N]}{\bar{A}}_{\ell-2} \overset{[N]}{\bar{B}}_{\ell-3} \overset{[1]}{x}_{\ell-3,e} +$$

$$\overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \overset{[N]}{\bar{B}}_{\ell-2} \overset{[1]}{x}_{\ell-2,e} +$$

$$\overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{B}}_{\ell-1} \overset{[1]}{x}_{\ell-1,e} +$$

$$\overset{[N]}{\bar{B}}_\ell \overset{[1]}{x}_{\ell,e}$$

Which can be written as

$$h_\ell =$$

$$(\overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \ldots \overset{[N]}{\bar{A}}_3 \overset{[N]}{\bar{A}}_2, \quad \overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \ldots \overset{[N]}{\bar{A}}_4 \overset{[N]}{\bar{A}}_3, \quad \ldots \quad \overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \overset{[N]}{\bar{A}}_{\ell-2}, \quad , \overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1}, \quad \overset{[N]}{\bar{A}}_\ell, \ \overset{[N]}{1})$$

$$(\overset{[N]}{\bar{B}}_1, \overset{[N]}{\bar{B}}_2, \ldots, \overset{[N]}{\bar{B}}_{\ell-3}, \overset{[N]}{\bar{B}}_{\ell-2}, \overset{[N]}{\bar{B}}_{\ell-1}, \overset{[N]}{\bar{B}}_\ell)$$

$$(\overset{[1]}{x}_{1,e}, \overset{[1]}{x}_{2,e}, \ldots, \overset{[1]}{x}_{\ell-3,e}, \overset{[1]}{x}_{\ell-2,e}, \overset{[1]}{x}_{\ell-1,e}, \overset{[1]}{x}_{\ell,e})$$

(doing element-wise multiplication of these three terms, than summing all the elements together.

Now consider the first term in

$$(\overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \ldots \overset{[N]}{\bar{A}}_3 \overset{[N]}{\bar{A}}_2, \quad \overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \ldots \overset{[N]}{\bar{A}}_4 \overset{[N]}{\bar{A}}_3, \quad \ldots \quad \overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \overset{[N]}{\bar{A}}_{\ell-2}, \quad , \overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1}, \quad \overset{[N]}{\bar{A}}_\ell, \ \overset{[N]}{1})$$

specifically this term:

$$\overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \ldots \overset{[N]}{\bar{A}}_3 \overset{[N]}{\bar{A}}_2$$

Proof

lets expand it out

$$= \exp(\overset{[1]}{\Delta}_{\ell,e} \overset{[N]}{A}[e]) \exp(\overset{[1]}{\Delta}_{\ell-1,e} \overset{[N]}{A}[e]) \ldots \exp(\overset{[1]}{\Delta}_{3,e} \overset{[N]}{A}[e]) \exp(\overset{[1]}{\Delta}_{2,e} \overset{[N]}{A}[e])$$

$$= \exp(\overset{[1]}{\Delta}_{\ell,e}\overset{[N]}{A}[e] + \overset{[1]}{\Delta}_{\ell-1,e}\overset{[N]}{A}[e] + \cdots + \overset{[1]}{\Delta}_{3,e}\overset{[N]}{A}[e] + \overset{[1]}{\Delta}_{2,e}\overset{[N]}{A}[e])$$

$$= \exp((\overset{[1]}{\Delta}_{\ell,e} + \overset{[1]}{\Delta}_{\ell-1,e} + \cdots + \overset{[1]}{\Delta}_{3,e} + \overset{[1]}{\Delta}_{2,e})\overset{[N]}{A}[e])$$

What is this inner delta term? Well

$$\overset{[1]}{\Delta}_{\ell,e} = \text{softplus}(\overset{[E]}{x}_{\ell} \cdot \overset{[E]}{W}_{\Delta}[:,e] + \overset{[1]}{B}_{\Delta}[e])$$

Lets call $\overset{[1]}{\delta}_{\ell,e} = \overset{[E]}{x}_{\ell} \cdot \overset{[E]}{W}_{\Delta}[:,e] + \overset{[1]}{B}_{\Delta}[e]$

Then, consider this simple case first:

$$\overset{[1]}{\Delta}_{1,e} + \overset{[1]}{\Delta}_{2,e}$$

$$= \text{softplus}(\overset{[1]}{\delta}_{1,e}) + \text{softplus}(\overset{[1]}{\delta}_{2,e})$$

Uselsss fun fact

Note,

$$\text{softplus}(x) = \log(1 + e^{x})$$

$$= \log\left(\frac{e^{-x} + 1}{e^{-x}}\right)$$

$$= \log(e^{-x} + 1) - \log(e^{-x})$$

$$= \log(1 + e^{-x}) + x$$

Means that

$$\overset{[1]}{\Delta}_{1,e} + \overset{[1]}{\Delta}_{2,e}$$

$$= \text{softplus}(\overset{[1]}{\delta}_{1,e}) + \text{softplus}(\overset{[1]}{\delta}_{2,e})$$

$$= \log\left(\left(1 + \exp(-\overset{[1]}{\delta}_{1,e})\right)\left(1 + \exp(-\overset{[1]}{\delta}_{2,e})\right)\right) + \overset{[1]}{\delta}_{1,e} + \overset{[1]}{\delta}_{2,e}$$

$$= \log\left(\exp(-\overset{[1]}{\delta}_{1,e} - \overset{[1]}{\delta}_{2,e}) + \exp(-\overset{[1]}{\delta}_{1,e}) + \exp(-\overset{[1]}{\delta}_{2,e}) + 1\right) + \overset{[1]}{\delta}_{1,e} + \overset{[1]}{\delta}_{2,e}$$

vs

Back to the proof

$$\text{softplus}(x) = \log(1 + e^{x})$$

$$\overset{[1]}{\Delta}_{1,e} + \overset{[1]}{\Delta}_{2,e}$$

$$= \text{softplus}(\overset{[1]}{\delta}_{1,e}) + \text{softplus}(\overset{[1]}{\delta}_{2,e})$$

$$= \log\Big(\big(1 + \exp(\overset{[1]}{\delta}_{1,e})\big)\big(1 + \exp(\overset{[1]}{\delta}_{2,e})\big) \Big)$$

$$= \log\Big(\exp(\overset{[1]}{\delta}_{1,e} + \overset{[1]}{\delta}_{2,e}) + \exp(\overset{[1]}{\delta}_{1,e}) + \exp(\overset{[1]}{\delta}_{2,e}) + 1 \Big)$$

I think I prefer this way of writing it because either way you gotta do all the subsets. Speaking of the subsets, we can rewrite this:

$$\log\Big(\exp(\overset{[1]}{\delta}_{1,e} + \overset{[1]}{\delta}_{2,e}) + \exp(\overset{[1]}{\delta}_{1,e}) + \exp(\overset{[1]}{\delta}_{2,e}) + 1 \Big)$$

$$= \log\Big(\sum_{S \subseteq \{1,2\}} \exp\big(\sum_{i \in S} \delta_{i,e} \big) \Big)$$

Where $S \subseteq \{1,2\}$ means loop through all subsets (including $\{1,2\}$ and the empty set, where empty set gives $\exp(0) = 1$)

Lets expand this out more: since $\overset{[1]}{\delta}_{\ell,e} = \overset{[E]}{x}_{\ell} \cdot \overset{[E]}{W}_{\Delta}[:,e] + \overset{[1]}{B}_{\Delta}[e]$

$$\log\Big(\sum_{S \subseteq \{1,2\}} \exp\big(\sum_{i \in S} \delta_{i,e} \big) \Big)$$

$$= \log\Big(\sum_{S \subseteq \{1,2\}} \exp\big(\sum_{j \in S} \overset{[E]}{x}_{j} \cdot \overset{[E]}{W}_{\Delta}[:,e] + \overset{[1]}{B}_{\Delta}[e] \big) \Big)$$

$$= \log\Big(\sum_{S \subseteq \{1,2\}} \exp\big(|S| \overset{[1]}{B}_{\Delta}[e] + \overset{[E]}{W}_{\Delta}[:,e] \cdot \sum_{i \in S} (\overset{[E]}{x}_{i}) \big) \Big)$$

$$= \log\Big(\sum_{S \subseteq \{1,2\}} \exp\big(|S| \overset{[1]}{B}_{\Delta}[e] \big) \exp\big(\overset{[E]}{W}_{\Delta}[:,e] \cdot \sum_{i \in S} (\overset{[E]}{x}_{i}) \big) \Big)$$

Thus

$$\overset{[1]}{\Delta}_{1,e} + \overset{[1]}{\Delta}_{2,e}$$

$$= \log\Big(\sum_{S \subseteq \{1,2\}} \exp\big(|S| \overset{[1]}{B}_{\Delta}[e] \big) \exp\big(\overset{[E]}{W}_{\Delta}[:,e] \cdot \sum_{i \in S} (\overset{[E]}{x}_{i}) \big) \Big)$$

Going back to

$$\overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \dots \overset{[N]}{\bar{A}}_3 \overset{[N]}{\bar{A}}_2$$

$$= \exp(\overset{[1]}{\Delta}_{\ell,e}\overset{[N]}{A}[e]) \exp(\overset{[1]}{\Delta}_{\ell-1,e}\overset{[N]}{A}[e]) \dots \exp(\overset{[1]}{\Delta}_{3,e}\overset{[N]}{A}[e]) \exp(\overset{[1]}{\Delta}_{2,e}\overset{[N]}{A}[e])$$

$$= \exp(\overset{[1]}{\Delta}_{\ell,e}\overset{[N]}{A}[e] + \overset{[1]}{\Delta}_{\ell-1,e}\overset{[N]}{A}[e] + \dots + \overset{[1]}{\Delta}_{3,e}\overset{[N]}{A}[e] + \overset{[1]}{\Delta}_{2,e}\overset{[N]}{A}[e])$$

$$= \exp((\overset{[1]}{\Delta}_{\ell,e} + \overset{[1]}{\Delta}_{\ell-1,e} + \dots + \overset{[1]}{\Delta}_{3,e} + \overset{[1]}{\Delta}_{2,e})\overset{[N]}{A}[e])$$

From above, which

$$\overset{[1]}{\Delta}_{\ell,e} + \overset{[1]}{\Delta}_{\ell-1,e} + \dots + \overset{[1]}{\Delta}_{3,e} + \overset{[1]}{\Delta}_{2,e}$$

$$= \text{softplus}(\overset{[1]}{\delta}_{\ell,e}) + \text{softplus}(\overset{[1]}{\delta}_{\ell-1,e}) + \dots + \text{softplus}(\overset{[1]}{\delta}_{3,e}) + \text{softplus}(\overset{[1]}{\delta}_{2,e})$$

$$= \log\Big(\sum_{S \subseteq \{\ell, \ell-1, \dots, 3, 2\}} \exp\big(\sum_{i \in S} \delta_{i,e}\big)\Big)$$

And since $\overset{[1]}{\delta}_{\ell,e} = \overset{[E]}{x}_\ell \cdot \overset{[E]}{W}_\Delta[:, e] + \overset{[1]}{B}_\Delta[e]$

$$\overset{[1]}{\Delta}_{\ell,e} + \overset{[1]}{\Delta}_{\ell-1,e} + \dots + \overset{[1]}{\Delta}_{3,e} + \overset{[1]}{\Delta}_{2,e}$$

$$= \log\Big(\sum_{S \subseteq \{\ell, \ell-1, \dots, 3, 2\}} \exp\big(\sum_{i \in S} \overset{[E]}{x}_i \cdot \overset{[E]}{W}_\Delta[:, e] + \overset{[1]}{B}_\Delta[e]\big)\Big)$$

$$= \log\Big(\sum_{S \subseteq \{\ell, \ell-1, \dots, 3, 2\}} \exp\big(|S|\overset{[1]}{B}_\Delta[e] + \overset{[E]}{W}_\Delta[:, e] \cdot \sum_{i \in S} \overset{[E]}{x}_i\big)\Big)$$

$$= \log\Big(\sum_{S \subseteq \{\ell, \ell-1, \dots, 3, 2\}} \exp\big(|S|\overset{[1]}{B}_\Delta[e]\big) \exp\big(\overset{[E]}{W}_\Delta[:, e] \cdot \sum_{i \in S} \overset{[E]}{x}_i\big)\Big)$$

Thus,

$$\overset{[N]}{\bar{A}}_\ell \overset{[N]}{\bar{A}}_{\ell-1} \dots \overset{[N]}{\bar{A}}_3 \overset{[N]}{\bar{A}}_2$$

$$= \exp(\log\Big(\sum_{S \subseteq \{\ell, \ell-1, \dots, 3, 2\}} \exp\big(|S|\overset{[1]}{B}_\Delta[e]\big) \exp\big(\overset{[E]}{W}_\Delta[:, e] \cdot \sum_{i \in S} \overset{[E]}{x}_i\big)\Big)\overset{[N]}{A}[e])$$

Now, note that $\exp(ab) = \exp(a)^b$ (assuming a and b are real). Thus

$$\exp(\log(a)b) = \exp(\log(a))^b = a^b$$

Giving us

$$\overset{[N]}{\bar{A}_\ell}\,\overset{[N]}{\bar{A}_{\ell-1}}\dots\overset{[N]}{\bar{A}_3}\,\overset{[N]}{\bar{A}_2}$$

$$= \mathrm{pow}\Big(\overbrace{\Big[\sum_{S\subseteq\{\ell,\ell-1,\dots,3,2\}} \exp\big(|S|\overset{[1]}{B_\Delta[e]}\big)\exp\big(\overset{[E]}{W_\Delta[:,e]}\cdot\sum_{i\in S}\overset{[E]}{x_i}\big)\Big]}^{[1]}, \overset{[N]}{A[e]}\Big)$$

Note that, as labeled, the big summation term results in a single float. So this is

$$r = \mathrm{pow}(\overset{[1]}{a}, \overset{[N]}{b})$$

where

$$r_i = \mathrm{pow}(a, b_i) = a^{b_i}$$