∘ $\mathrm{GeLU}(x) = x \cdot P(X \geq x)$

• $\frac{\partial}{\partial x} \operatorname{erf}(\frac{x}{\sqrt{2}}) = \frac{\partial}{\partial x}\left(\frac{2}{\sqrt{\pi}} \int_0^{\frac{x}{\sqrt{2}}} e^{-t^2} dt\right)$

$= \frac{\partial}{\partial x}\left(\frac{2}{\sqrt{\pi}} \int_0^{x} e^{-\frac{s^2}{2}} d\frac{s}{\sqrt{2}}\right)$

$= \frac{1}{\sqrt{2\pi}} \int_0^{x} e^{-\frac{s^2}{2}} ds = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$\frac{\partial A^T}{\partial A} =$

• $P(X \geq x) = \frac{1}{2}(1 + \operatorname{erf}(\frac{x}{\sqrt{2}}))$

$\frac{\partial}{\partial x} \mathrm{GeLU}(x) = P(X \geq x) + x \frac{\partial}{\partial x} P(X \geq x)$

$= \operatorname{erf}(x) + \frac{1}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}}$

∘ Leaky ReLU $(x) = L(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha x & \text{if } x < 0 \end{cases}$

for $x > 0$, $\frac{\partial L}{\partial x} = 1$

for $x < 0$, $\frac{\partial L}{\partial x} = \alpha$

for $x = 0$, $\lim_{h^+ \to 0} \frac{L(x+h) - L(x)}{h} \neq \lim_{h^- \to 0} \frac{L(x+h) - L(x)}{h}$

It's derivative does not exist. Here we assign it as 1.

$$\frac{\partial L}{\partial x} = \begin{cases} 1 & \text{if } x \geq 0 \\ \alpha & \text{if } x < 0 \end{cases}$$

0 Prove $\nabla_A L = \frac{\partial L}{\partial C} B^T$

If A, B, C are 2nd-order tensors:

$$\nabla_A L = \frac{\partial L}{\partial A} \qquad (\nabla_A L)_{pq} = \frac{\partial L}{\partial A_{pq}} = \sum_i \sum_k \frac{\partial L}{\partial C_{ik}} \frac{\partial C_{ik}}{\partial A_{pq}}$$

$$\frac{\partial C_{ik}}{\partial A_{pq}} = \frac{\partial \sum_j A_{ij} B_{jk}}{\partial A_{pq}}$$

if $i = p$, $j = q$ $\frac{\partial A_{ij}}{\partial A_{pq}} = 1$, otherwise 0

$$\Rightarrow \frac{\partial L}{\partial A_{pq}} = \sum_k \left(\sum_i \frac{\partial L}{\partial C_{ik}} \frac{\partial \sum_j A_{ij} B_{jk}}{\partial A_{pq}} \right)$$

$$\overset{i=p,\, j=q}{\underset{\text{others } 0}{=\!=\!=}} \sum_k \left(\frac{\partial L}{\partial C_{pk}} B_{qk} \right) = \sum_k \left(\frac{\partial L}{\partial C_{pk}} B^T_{kq} \right)$$

$$= \left\{ \frac{\partial L}{\partial C} B^T \right\}_{pq}$$

Prove $\nabla_B L = A^T \frac{\partial L}{\partial C}$

$$(\nabla_B L)_{pq} = \frac{\partial L}{\partial B_{pq}} = \sum_i \sum_k \frac{\partial L}{\partial C_{ik}} \frac{\partial C_{ik}}{\partial B_{pq}}$$

$$= \sum_i \sum_k \frac{\partial L}{\partial C_{ik}} \frac{\partial \sum_j A_{ij} B_{jk}}{\partial B_{pq}}$$

$$\underset{\substack{j=p \\ k=q}}{=\!=\!=} \sum_i \frac{\partial L}{\partial C_{iq}} A_{ip} = \sum_i A^T_{pi} \frac{\partial L}{\partial C_{iq}}$$

$$= \left(A^T \frac{\partial L}{\partial C} \right)_{pq}$$

∘ More Generally

Suppose u, u' are arbitrary strings in the indices string sets of A e.g. $\mathbb{R}^{m\times n\times o}$ $\{123, 343, \dots\}$

Denote the set as $\mathcal{U}$ $\{12, 34, 45 \dots\}$ e.g. $\mathbb{R}^{m\times n}$

v, v' are arbitrary strings in the indices string sets of B

Denote the set as V

w - - - - - - - - C

Denote the set as W

~~$\mathcal{U}$~~

Additionally, we define a indices string sets J, which collects all the indices as string which are in $\forall \hat{u} \in \mathcal{U}$ and $\forall \hat{v} \in V$:

$$J := \{ j \mid j \subseteq u, j \subseteq v, u \in \mathcal{U}, v \in V \}$$

Similarly define $I := \{ i \mid i \subseteq u, i \not\subseteq v, u \in \mathcal{U}, v \in V \}$,

$$K := \{ k \mid k \not\subseteq u, k \subseteq v, u \in \mathcal{U}, v \in V \}$$

Apparently
$$\mathcal{U} = \{ i \& j \mid i \in I, j \in J \}$$
$$V = \{ v = j \& k \mid i \in I, j \in J \}$$
$$W = \{ w = i \& k \mid i \in J, k \in K \} \quad \Big(\sum_{j\in J} A_{ij} B_{jk} = C_{ik} \Big)$$

Then for u-th element of A, let $u = i^* \& j^*$

$$\frac{\partial L}{\partial A_u} \underset{\text{chain rule}}{=} \sum_{w\in W} \frac{\partial L}{\partial C_w} \frac{\partial C_w}{\partial A_u}$$

for any $\hat{w} \in W$, we have

$$C_{\hat{w}} = C_{\hat{i}\hat{k}} = \sum_{j\in J} A_{\hat{i}j} B_{j\hat{k}} \qquad \hat{w} = \hat{i} \& \hat{k}, \hat{i} \in I, \hat{k} \in K$$

$$\Rightarrow \frac{\partial L}{\partial A_u} = \sum_{i\in I, k\in K} \frac{\partial L}{\partial C_{ik}} \frac{\partial \left(\sum_{j\in J} A_{ij} B_{jk}\right)}{\partial A_{i^*j^*}}$$

Now we look at $\frac{\partial L}{\partial C_{ik}} \frac{\partial}{\partial A_{i^*j^*}} \left(\sum_{j\in J} A_{ij} B_{jk} \right)$

$$\frac{\partial (A_{ij} B_{jk})}{\partial A_{i^*j^*}} = \begin{cases} B_{jk} & \text{if } i^*=i, j^*=j \\ 0 & \text{otherwise} \end{cases} = \begin{cases} B_{j^*k} & \text{if } i^*=i, j^*=j \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial L}{\partial A_u} = \sum_k \frac{\partial L}{\partial C_{i^*k}} B_{j^*k} = \left(\sum_k \frac{\partial L}{\partial C_{i^*k}} B_{j^*k} \right)_{i^*j^*}$$

since $u = i^*j^*$ is arbitrarily chosen $\Rightarrow$ $\text{einsum}(\underbrace{j\&k}_{v}, \underbrace{i\&k}_{w} \to \underbrace{i\&j}_{u}, B, \frac{\partial L}{\partial C}) = \frac{\partial L}{\partial A}$

$$= \text{einsum}(v, w \to u, B, \frac{\partial L}{\partial C}) \qquad \square$$

I won't do it again for B since it's not worth it.

Important: $$\frac{\partial A_{ij} B_{jk}}{\partial B_{j^*k^*}} = \begin{cases} A_{ij^*} & j^* = j,\ k^* = k \\ 0 & \text{otherwise} \end{cases}$$