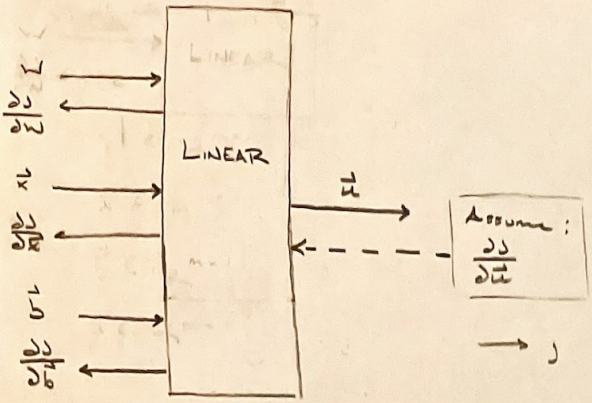
PAIGE ROYNEK
CS 3450 LAB 5 - DERIVING GRADIENTS

1. LINEAR LAYER



Forward Propagation:

$$\vec{u} = W\vec{x} + \vec{b}$$

where:

- $\rightarrow W \in \mathbb{R}^{m \times n}$
- $\rightarrow \vec{x} \in \mathbb{R}^{n \times 1}$
- $\rightarrow \vec{b} \in \mathbb{R}^{m \times 1}$
- $\rightarrow \vec{u} \in \mathbb{R}^{m \times 1}$

Backpropagation:

$$\frac{\partial J}{\partial W} = \text{prod}\left(\frac{\partial J}{\partial \vec{u}}, \frac{\partial \vec{u}}{\partial W}\right) = \boxed{\frac{\partial J}{\partial \vec{u}} \cdot \vec{x}^T}$$

$$\frac{\partial J}{\partial \vec{b}} = \text{prod}\left(\frac{\partial J}{\partial \vec{u}}, \frac{\partial \vec{u}}{\partial \vec{b}}\right) = \frac{\partial J}{\partial \vec{u}} \cdot 1 = \boxed{\frac{\partial J}{\partial \vec{u}}}$$

$$\frac{\partial J}{\partial \vec{x}} = \text{prod}\left(\frac{\partial J}{\partial \vec{u}}, \frac{\partial \vec{u}}{\partial \vec{x}}\right) = \boxed{\frac{\partial J}{\partial \vec{u}} \cdot W^T}$$

NOTE: $\vec{u} = W\vec{x} + \vec{b}$ (expanded)

$$\vec{u} = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \vdots & \vdots \\ w_{m1} & \cdots & w_{mn} \end{bmatrix}_{m \times n} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{n \times 1} + \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}_{m \times 1} = \overbrace{\begin{bmatrix} w_{11}x_1 + \cdots + w_{1n}x_n + b_1 \\ \vdots \\ w_{m1}x_1 + \cdots + w_{mn}x_n + b_m \end{bmatrix}_{m \times 1}}^{\vec{u}} \begin{matrix} \Rightarrow u_1 \\ \\ \Rightarrow u_m \end{matrix}$$

element-wise: $\frac{\partial \vec{u}}{\partial W}$

$$\frac{\partial \vec{u}}{\partial W} = \begin{bmatrix} \frac{\partial \vec{u}}{\partial w_{11}} & \cdots & \frac{\partial \vec{u}}{\partial w_{1n}} \\ \vdots & \vdots & \vdots \\ \frac{\partial \vec{u}}{\partial w_{m1}} & \cdots & \frac{\partial \vec{u}}{\partial w_{mn}} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \vec{x}^T$$

where:

$$\frac{\partial \vec{u}}{\partial w_{11}} = \begin{bmatrix} \frac{\partial u_1}{\partial w_{11}} \\ \vdots \\ \frac{\partial u_m}{\partial w_{11}} \end{bmatrix}_{m \times 1} = \begin{bmatrix} x_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{m \times 1}$$

$$\frac{\partial \vec{u}}{\partial w_{11}} = x_1$$

element-wise: $\frac{\partial \vec{u}}{\partial \vec{b}}$

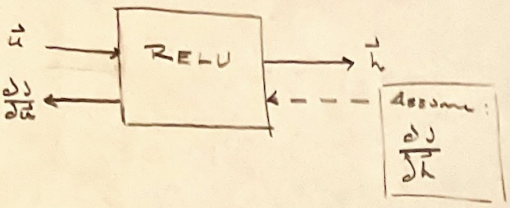
$$\frac{\partial \vec{u}}{\partial \vec{b}} = \begin{bmatrix} \frac{\partial \vec{u}}{\partial b_1} \\ \vdots \\ \frac{\partial \vec{u}}{\partial b_m} \end{bmatrix}_{m \times 1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{m \times 1} = 1$$

where:

$$\frac{\partial \vec{u}}{\partial b_1} = \begin{bmatrix} \frac{\partial u_1}{\partial b_1} \\ \vdots \\ \frac{\partial u_m}{\partial b_1} \end{bmatrix}_{m \times 1} \Rightarrow \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \end{bmatrix}_{m \times 1} = 1$$

element-wise $\frac{\partial \vec{u}}{\partial \vec{x}}$ ON PAGE 6 $\Longrightarrow$

(2.) ReLU LAYER



Forward Propagation:

$$\vec{h} = \text{ReLU}(\vec{u})$$

where:

$\vec{h} \in \mathbb{R}^{m \times 1}$

$\vec{u} \in \mathbb{R}^{m \times 1}$

NOTE:

$$\text{ReLU}(u_i) = \begin{cases} u_i, & \text{if } u_i > 0 \\ 0, & \text{otherwise} \end{cases}$$

Backpropagation:

$$\frac{\partial J}{\partial \vec{u}} = \text{prod}\left(\frac{\partial J}{\partial \vec{h}}, \frac{\partial \vec{h}}{\partial \vec{u}}\right)$$

$$= \frac{\partial J}{\partial \vec{h}} \odot \text{ReLU}'(\vec{u})$$

$$\text{ReLU}'(\vec{u}) = \begin{cases} 1, & \text{if } u_i > 0 \\ 0, & \text{otherwise} \end{cases} = \vec{1}_{u_i > 0}$$

$$\boxed{\frac{\partial J}{\partial \vec{u}} = \frac{\partial J}{\partial \vec{h}} \odot \vec{1}_{u_i > 0}}$$

vector

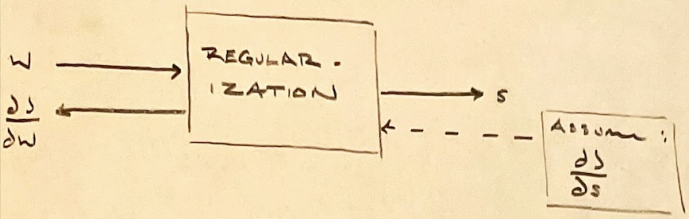
element-wise: $\frac{\partial \vec{h}}{\partial \vec{u}}$

$$\frac{\partial \vec{h}}{\partial \vec{u}} = \begin{bmatrix} \frac{\partial h_1}{\partial \vec{u}} \\ \vdots \\ \frac{\partial h_m}{\partial \vec{u}} \end{bmatrix}_{m \times 1} = \vec{1}_{u_i > 0}$$

where:

$$\frac{\partial h_1}{\partial \vec{u}} = \begin{bmatrix} \frac{\partial h_1}{\partial u_1} \\ \vdots \\ \frac{\partial h_1}{\partial u_m} \end{bmatrix}_{m \times 1} = \vec{1}_{u_i > 0}$$

③ REGULARIZATION LAYER



Forward Propagation:

$$s = \frac{\lambda}{2}\|W\|_F^2$$

$$= \frac{\lambda}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}(w_{ij})^2$$

where:

$W \in \mathbb{R}^{m \times n}$

$s \in \mathbb{R}^1$ (scalar)

$\lambda \in \mathbb{R}^1$ (scalar, hyperparameter)

Backpropagation:

$$\frac{\partial J}{\partial W} = \text{prod}\left(\frac{\partial J}{\partial s}, \frac{\partial s}{\partial W}\right)$$

$$= \boxed{\frac{\partial J}{\partial s} \cdot \lambda W}$$

element-wise: $\frac{\partial s}{\partial W}$

$$\frac{\partial s}{\partial W} = \begin{bmatrix} \frac{\partial s}{\partial w_{11}} & \cdots & \frac{\partial s}{\partial w_{1n}} \\ \vdots & \vdots & \vdots \\ \frac{\partial s}{\partial w_{m1}} & \cdots & \frac{\partial s}{\partial w_{mn}} \end{bmatrix}_{m \times n}$$

where:

$$s = \frac{\lambda}{2} \cdot \left[w_{11}^2 + \cdots + w_{1n}^2 + \cdots + w_{m1}^2 + \cdots + w_{mn}^2\right]$$

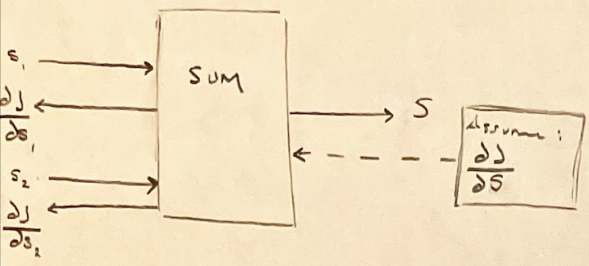
then:

$$\frac{\partial s}{\partial W} = \begin{bmatrix} \lambda w_{11} & \cdots & \lambda w_{1n} \\ \vdots & \vdots & \vdots \\ \lambda w_{m1} & \cdots & \lambda w_{mn} \end{bmatrix}_{m \times n} = \lambda W$$

where:

$$\frac{\partial s}{\partial w_{11}} = \lambda w_{11}$$

4. SUM LAYER



Forward Propagation:

$$S = s_1 + s_2$$

where:

$S, s_1, s_2 \in \mathbb{R}^1$ (scalar)

Backpropagation:

$$\frac{\partial J}{\partial s_1} = \text{prod}\left(\frac{\partial J}{\partial S}, \frac{\partial S}{\partial s_1}\right)$$

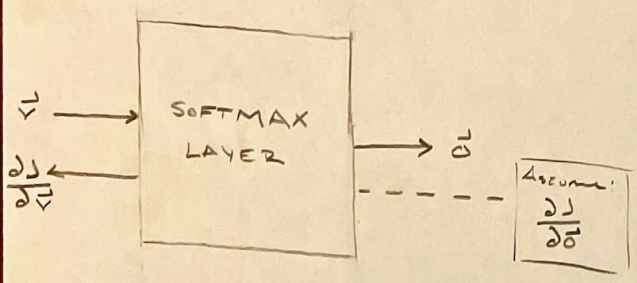
$$= \frac{\partial J}{\partial S} \cdot 1$$

$$\boxed{\frac{\partial J}{\partial s_1} = \frac{\partial J}{\partial S}}$$

where:

$$\frac{\partial S}{\partial s_1} = \frac{\partial}{\partial s_1}\left[s_1 + \cancel{s_2}^{\,0}\right] = 1$$

5. SOFTMAX LAYER



Forward Propagation:

$$\vec{o} = \text{softmax}(\vec{v})$$

$$= \frac{e^{(v_i)}}{\sum_{j=1}^{m} e^{v_j}}$$

where:

$\vec{o} \in \mathbb{R}^{m \times 1}$

$\vec{v} \in \mathbb{R}^{m \times 1}$

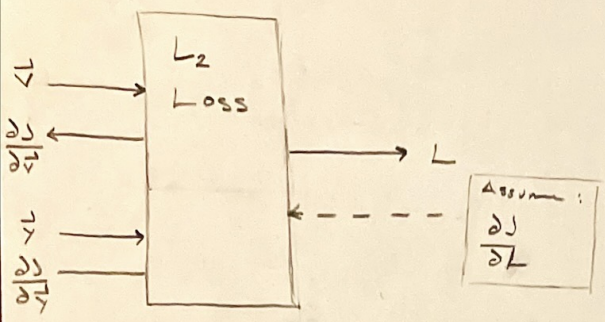
Backpropagation:

$$\frac{\partial J}{\partial \vec{v}} = \text{prod}\left(\frac{\partial J}{\partial \vec{o}}, \frac{\partial \vec{o}}{\partial \vec{v}}\right)$$

$$= \boxed{\frac{\partial J}{\partial \vec{o}} \cdot (\vec{v} - \vec{y})}$$

from lib doc
DON'T HAVE
TO DERIVE

6. L_2 LOSS LAYER



Forward Propagation:

$$L = \frac{1}{2} \| \vec{v} - \vec{y} \|_2^2$$

$$= \frac{1}{2} \left[(v_1 - y_1)^2 + \cdots + (v_m - y_m)^2 \right]$$

where:

$\vec{v} \in \mathbb{R}^{m \times 1}$

$\vec{y} \in \mathbb{R}^{m \times 1}$

$L \in \mathbb{R}^1$ (scalar)

Back propagation:

$$\frac{\partial J}{\partial \vec{v}} = \text{prod}\left(\frac{\partial J}{\partial L}, \frac{\partial L}{\partial \vec{v}} \right)$$

$$\boxed{\frac{\partial J}{\partial \vec{v}} = \frac{\partial J}{\partial L} \cdot (\vec{v} - \vec{y})}$$

expanded: $\frac{\partial L}{\partial \vec{v}}$

$$\frac{\partial L}{\partial \vec{v}} = \begin{bmatrix} \frac{\partial L}{\partial v_1} \\ \vdots \\ \frac{\partial L}{\partial v_m} \end{bmatrix}_{m \times 1} = \begin{bmatrix} v_1 - y_1 \\ \vdots \\ v_m - y_m \end{bmatrix} = \underline{\vec{v} - \vec{y}}$$

Note:

Let $g(v_i) = v_i - y_i$

Let $f(g) = \frac{1}{2} g^2(v_i)$

$$\frac{\partial f}{\partial v_i} = \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial v_i}$$

$$= \left[\frac{1}{2} \cdot 2 \; g(v_i) \right] \cdot 1$$

$$= g(v_i) = \underline{(v_i - y_i)}$$

ELEMENTWISE EQNS: CONT.

LINEAR LAYER:

<u>element-wise</u>: $\frac{\partial \vec{u}}{\partial \vec{x}}$

$$\frac{\partial \vec{u}}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial \vec{u}}{\partial x_1} \\ \vdots \\ \frac{\partial \vec{u}}{\partial x_n} \end{bmatrix}_{n \times 1} = \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \cdots & \frac{\partial u_m}{\partial x_1} \\ \vdots & \vdots & \vdots \\ \frac{\partial u_1}{\partial x_n} & \cdots & \frac{\partial u_m}{\partial x_n} \end{bmatrix}_{n \times m} = \begin{bmatrix} w_{11} & \cdots & w_{m1} \\ \vdots & \vdots & \vdots \\ w_{1n} & \cdots & w_{nm} \end{bmatrix}_{n \times m} = \boxed{W^T}$$