RIGE ROSYNEK
CS 3450 021
LAB 6 - IMPLEMENTING FORWARD PROPAGATION

FORWARD PROPAGATION (BY LAYER TYPE)

1. LINEAR LAYER

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

where:

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

$\vec{x} \in \mathbb{R}^{n \times 1}$

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

EXAMPLE:

$$W = \begin{bmatrix} 1 & 5 & 7 \\ 3 & 2 & 4 \end{bmatrix}_{2\times 3}, \quad \vec{x} = \begin{bmatrix} 2 \\ 10 \\ 5 \end{bmatrix}_{3\times 1}, \quad \vec{b} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}_{2\times 1}$$

$$\vec{u} = \begin{bmatrix} 1 & 5 & 7 \\ 3 & 2 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 10 \\ 5 \end{bmatrix} + \begin{bmatrix} 4 \\ 8 \end{bmatrix}$$

$$= \begin{bmatrix} (1)(2) + (5)(10) + (7)(5) \\ (3)(2) + (2)(5) + (4)(5) \end{bmatrix} + \begin{bmatrix} 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 97 \\ 46 \end{bmatrix} + \begin{bmatrix} 4 \\ 8 \end{bmatrix}$$

$$\boxed{\vec{u} = \begin{bmatrix} 91 \\ 54 \end{bmatrix}_{2\times 1}}$$

2. ReLU LAYER

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

where:

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

NOTE:

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

EXAMPLE:

$$\vec{u} = \begin{bmatrix} -10 \\ 4 \\ 0 \end{bmatrix}_{3\times 1}$$

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

$$= ReLU\left(\begin{bmatrix} -10 \\ 4 \\ 0 \end{bmatrix}\right) \rightarrow \begin{array}{l} ReLU(-10) = 0 \\ ReLU(4) = 4 \\ ReLU(0) = 0 \end{array}$$

$$\boxed{\vec{h} = \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix}_{3\times 1}}$$

3. REGULARIZATION LAYER

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

where:

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

$\lambda, s \in \mathbb{R}^1$

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

EXAMPLE:

$$W = \begin{bmatrix} 1 & 5 & 7 \\ 3 & 2 & 4 \end{bmatrix}_{2 \times 3}, \quad \lambda = 0.1$$

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

$$= \frac{0.1}{2} \left((1)^2 + (5)^2 + (7)^2 + (3)^2 + (2)^2 + (4)^2 \right)$$

$$= \frac{0.1}{2} (104) = 0.05(104)$$

$$\boxed{s = 5.2}$$

4. SUM LAYER

$$s = s_1 + s_2$$

where:

$s, s_1, s_2 \in \mathbb{R}$

EXAMPLE:

$s_1 = 50, \quad s_2 = 5$

$$s = 50 + 5$$

$$\boxed{s = 55}$$

5. L_2 Loss Layer (MSE)

$$L = \frac{1}{2} \left\| \vec{o} - \vec{y} \right\|_2^2$$

$$= \frac{1}{2} \left[(o_1 - y_1)^2 + \dots + (o_m - y_m)^2 \right]$$

where:

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

$L \in \mathbb{R}$

Example:

$$\vec{o} = \begin{bmatrix} 10 \\ 5 \end{bmatrix}_{2\times1}, \quad \vec{y} = \begin{bmatrix} 11 \\ 5 \end{bmatrix}_{2\times1}$$

$$L = \frac{1}{2} \cdot \left\| \begin{bmatrix} 10 \\ 5 \end{bmatrix} - \begin{bmatrix} 11 \\ 5 \end{bmatrix} \right\|_2^2$$

$$= \frac{1}{2} \left[(10-11)^2 + (5-5)^2 \right] = \frac{1}{2}(1+0)$$

$$\boxed{L = 0.5}$$

4. Softmax Layer

softmax: $\vec{o}$

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

where:

$\vec{o}, \vec{v} \in \mathbb{R}^{d\times1}$

cross-entropy loss: L

$$L = CE(\vec{o}, \vec{y})$$

$$= -\frac{1}{N} \sum_{j=1}^{N} \left[y_j \cdot \log(o_j) \right]$$

where:

$\vec{o}, \vec{y} \in \mathbb{R}^{d\times1}$

$L \in \mathbb{R}$

NOTE:

$$\text{softmax}(v_i) = \frac{e^{(v_i - \max(\vec{v}))}}{\sum_{j=1}^{d} e^{(v_j - \max(\vec{v}))}}$$

EXAMPLE
NEXT PAGE ⇒

EXAMPLE: softmax layer

$$\vec{v} = \begin{bmatrix} 10 \\ 5 \end{bmatrix}_{2\times 1}$$

softmax:

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

$$= \text{softmax}\left(\begin{bmatrix} 10 \\ 5 \end{bmatrix}\right) \Rightarrow$$

$$\text{softmax}(\overset{v_1}{10}) = \frac{e^{10-10}}{e^{10-10} + e^{5-10}} = \frac{1}{1+e^{-5}} = \underline{0.993}$$

$$\text{softmax}(\overset{v_2}{5}) = \frac{e^{5-10}}{e^{10-10} + e^{5-10}} = \frac{e^{-5}}{1+e^{-5}} = \underline{0.007}$$

$$\boxed{\vec{o} = \begin{bmatrix} 0.993 \\ 0.007 \end{bmatrix}_{2\times 1}}$$

check:

→ should sum to 1

$0.993 + 0.007 = 1$ ✓

cross-entropy loss:

$$\vec{o} = \begin{bmatrix} 0.993 \\ 0.007 \end{bmatrix}_{2\times 1}, \quad \vec{y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{2\times 1}$$

$$L = -\frac{1}{2} \sum_{j=1}^{2} \left(y_j \cdot \log(o_j) \right)$$

$$= -\frac{1}{2} \left[(1) \cdot \log(0.993) + (0) \cdot \log(0.007) \right]$$

$$= -\frac{1}{2} \cdot \left[-0.00672 \right]$$

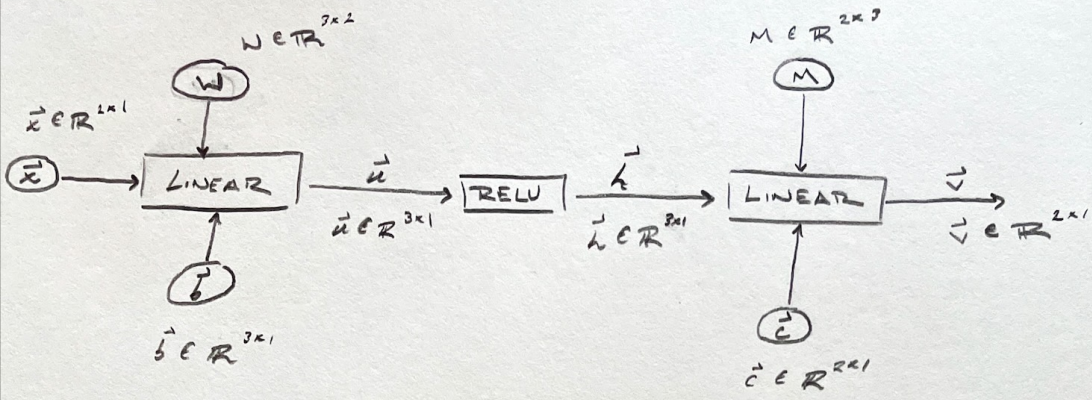
$$\boxed{L = 0.003358}$$

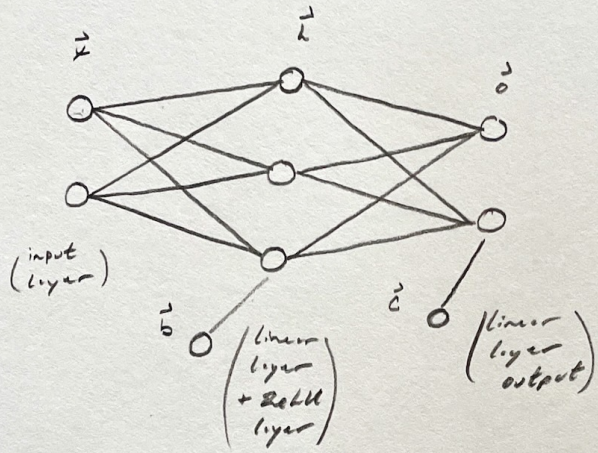
NOTE:

log(x) is ln(x)

FORWARD PROPAGATION
FULL NETWORK

NETWORK:





Given:

$$\vec{x} = \begin{bmatrix} 10 \\ 5 \end{bmatrix}_{2\times 1}, \quad W = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 7 & 3 \end{bmatrix}_{3\times 2}, \quad \vec{b} = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}_{3\times 1}, \quad \vec{c} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}_{2\times 1}$$

$$M = \begin{bmatrix} 4 & 1 & 2 \\ 3 & 6 & 1 \end{bmatrix}_{2\times 3}$$

1. LINEAR LAYER (hidden layer)

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

$$\vec{z} = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 7 & 3 \end{bmatrix} \begin{bmatrix} 10 \\ 5 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} (1)(10) + (2)(5) \\ (2)(10) + (5)(5) \\ (7)(10) + (3)(5) \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 20 \\ 45 \\ 85 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$$

$$\boxed{\vec{z} = \begin{bmatrix} 24 \\ 47 \\ 86 \end{bmatrix}}$$

ReLU:

$$\vec{h} = \mathrm{ReLU}(\vec{z}) = \mathrm{ReLU}\left(\begin{bmatrix} 24 \\ 47 \\ 86 \end{bmatrix}\right) \Rightarrow \begin{array}{l} \mathrm{ReLU}(24) = 24 \\ \mathrm{ReLU}(47) = 47 \\ \mathrm{ReLU}(86) = 86 \end{array}$$

$$\boxed{\vec{h} = \begin{bmatrix} 24 \\ 47 \\ 86 \end{bmatrix}}$$

2. LINEAR LAYER (output layer)

$$\vec{o} = M\vec{h} + \vec{c}$$

$$= \begin{bmatrix} 4 & 1 & 2 \\ 3 & 6 & 1 \end{bmatrix} \begin{bmatrix} 24 \\ 47 \\ 86 \end{bmatrix} + \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} (4)(24) + (1)(47) + (2)(86) \\ (3)(24) + (6)(47) + (1)(86) \end{bmatrix} + \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

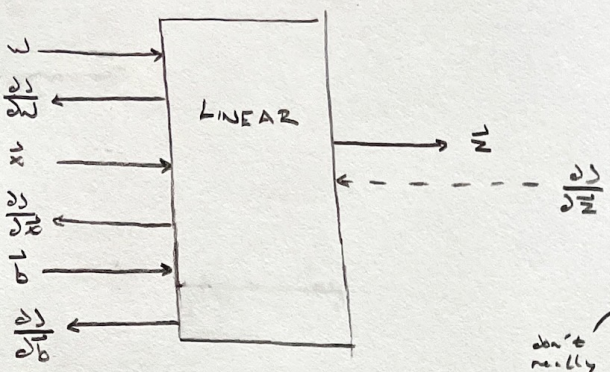
$$= \begin{bmatrix} 315 \\ 440 \end{bmatrix} + \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

$$\boxed{\vec{o} = \begin{bmatrix} 320 \\ 441 \end{bmatrix}_{2\times 1}}$$

BACKPROPAGATION BY LAYER TYPE

① LINEAR LAYER

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



have: $\frac{\partial J}{\partial \vec{z}}$

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

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

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

don't really need

EXAMPLE:

$$\vec{x} = \begin{bmatrix} 10 \\ 5 \end{bmatrix}_{2\times 1}, \quad W = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 7 & 3 \end{bmatrix}_{3\times 2}, \quad \vec{b} = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}_{3\times 1}$$

Backprop:

$$\boxed{\frac{\partial J}{\partial W} = \frac{\partial J}{\partial \vec{z}} \cdot \begin{bmatrix} 10 & 5 \end{bmatrix}}$$

$$\boxed{\frac{\partial J}{\partial \vec{x}} = \frac{\partial J}{\partial \vec{z}} \cdot \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 7 & 3 \end{bmatrix}}$$

$$\boxed{\frac{\partial J}{\partial \vec{b}} = \frac{\partial J}{\partial \vec{z}} \cdot 1}$$

step (1): Gradient descent $\quad \alpha = 0.01$

$$W^* = W - \alpha \frac{\partial J}{\partial W}$$

$$\boxed{W^* = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 7 & 3 \end{bmatrix} - 0.01\left(\frac{\partial J}{\partial \vec{z}} \cdot \begin{bmatrix} 10 & 5 \end{bmatrix}\right)}$$

$$b^* = \vec{b} - \alpha \frac{\partial J}{\partial \vec{b}}$$

$$\boxed{b^* = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} - 0.01\left(\frac{\partial J}{\partial \vec{z}}\right)}$$

② RELU LAYER

Have: $\frac{\partial J}{\partial \vec{h}}$

$$\frac{\partial J}{\partial \vec{z}} = \text{prod}\left(\frac{\partial J}{\partial \vec{h}}, \frac{\partial \vec{h}}{\partial \vec{z}}\right) = \boxed{\frac{\partial J}{\partial \vec{h}} \odot \text{ReLU}'(\vec{z})}$$

where:

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

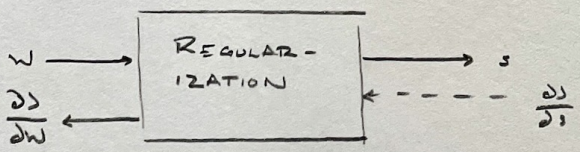
EXAMPLE:

$$\vec{z} = \begin{bmatrix} -4 \\ 10 \\ 5 \end{bmatrix}_{3\times 1}$$

$$\frac{\partial J}{\partial \vec{z}} = \frac{\partial J}{\partial \vec{h}} \odot \text{ReLU}'(\vec{z}) = \frac{\partial J}{\partial \vec{h}} \odot \text{ReLU}'\left(\begin{bmatrix} -4 \\ 10 \\ 5 \end{bmatrix}\right)$$

$$\boxed{\frac{\partial J}{\partial \vec{z}} = \frac{\partial J}{\partial \vec{h}} \odot \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}}$$

③ REGULARIZATION LAYER



Have: $\frac{\partial J}{\partial s}$

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

EXAMPLE:

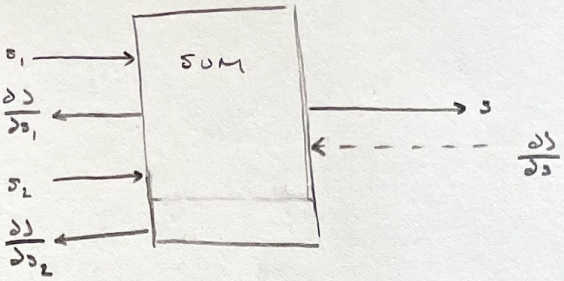
$$W = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 7 & 3 \end{bmatrix}_{3\times 2}, \quad \lambda = 0.1$$

$$\frac{\partial J}{\partial W} = \frac{\partial J}{\partial s} \cdot (0.1) \cdot \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 7 & 3 \end{bmatrix}$$

$$\boxed{\frac{\partial J}{\partial W} = \frac{\partial J}{\partial s} \cdot \begin{bmatrix} 0.1 & 0.2 \\ 0.2 & 0.5 \\ 0.7 & 0.3 \end{bmatrix}_{3\times 2}}$$

4. SUM LAYER

$s = s_1 + s_2$



h.n: $\frac{\partial J}{\partial s}$

$$\frac{\partial J}{\partial s_1} = \text{prod}\left(\frac{\partial J}{\partial s}, \frac{\partial s}{\partial s_1}\right) = \boxed{\frac{\partial J}{\partial s} \cdot 1}$$

$$\frac{\partial J}{\partial s_2} = \text{prod}\left(\frac{\partial J}{\partial s}, \frac{\partial s}{\partial s_2}\right) = \boxed{\frac{\partial J}{\partial s} \cdot 1}$$

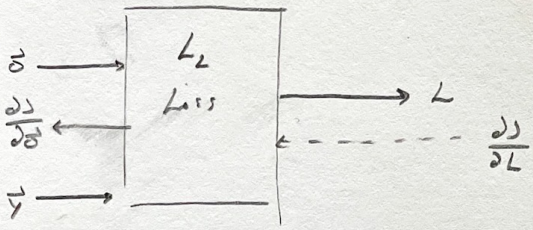
EXAMPLE:

$s_1 = 10, \quad s_2 = 50$

$$\boxed{\frac{\partial J}{\partial s_1} = \frac{\partial J}{\partial s} \cdot 1}$$

$$\boxed{\frac{\partial J}{\partial s_2} = \frac{\partial J}{\partial s} \cdot 1}$$

5. L_2 LOSS LAYER (MSE)



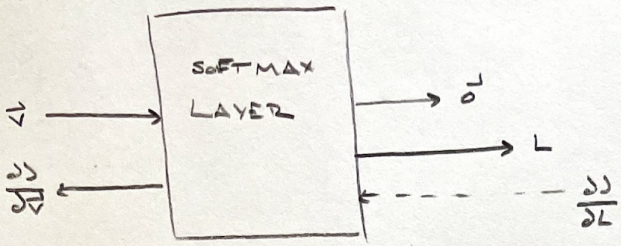
h.n: $\frac{\partial J}{\partial L}$

$$\frac{\partial J}{\partial \vec{o}} = \text{prod}\left(\frac{\partial J}{\partial L}, \frac{\partial L}{\partial \vec{o}}\right) = \boxed{\frac{\partial J}{\partial L} \cdot (\vec{o} - \vec{y})}$$

EXAMPLE:

$$\vec{o} = \begin{bmatrix} 9 \\ 2 \end{bmatrix}_{2\times 1}, \quad \vec{y} = \begin{bmatrix} 10 \\ 2 \end{bmatrix}_{2\times 1}$$

$$\frac{\partial J}{\partial \vec{o}} = \frac{\partial J}{\partial L} \cdot \left(\begin{bmatrix} 9 \\ 2 \end{bmatrix} - \begin{bmatrix} 10 \\ 2 \end{bmatrix}\right)$$

$$\boxed{\frac{\partial J}{\partial \vec{o}} = \frac{\partial J}{\partial L} \cdot \begin{bmatrix} -1 \\ 0 \end{bmatrix}}$$

(4.) SOFTMAX LAYER: (w/ cross-entropy)