Backpropagation [illegible]

[illegible]

Objective Function:

$\Rightarrow J = L + s$ where: $\frac{\partial J}{\partial L} = 1$, $\frac{\partial J}{\partial s} = 1$

① Linear Layer

$\Rightarrow \vec{z} = W\vec{x} + \vec{b}$ (forward propagation)

Backpropagation:
→ have: $\frac{\partial J}{\partial \vec{z}}$
→ goal: $\frac{\partial J}{\partial W}$, $\frac{\partial J}{\partial \vec{x}}$, $\frac{\partial J}{\partial \vec{b}}$ (optional)

i) [illegible]

ii) [illegible]

iii) [illegible]

EXAMPLE:

Given: [illegible]

Assume: [illegible]

then: [illegible]

[illegible]

$$\boxed{b^* = \begin{bmatrix} 3.9 \\ 7.9 \end{bmatrix}_{2\times1}}$$

② ReLU Layer

$\Rightarrow \vec{h} = relu(\vec{v})$ (forward propagation)

where:
$$relu(v_i) = \begin{cases} v_i, & \text{if } v_i > 0 \\ 0, & \text{otherwise} \end{cases}$$

Backpropagation:
→ have: $\frac{\partial J}{\partial \vec{h}}$
→ goal: $\frac{\partial J}{\partial \vec{v}}$

$$\frac{\partial J}{\partial \vec{v}} = prod\left(\frac{\partial J}{\partial \vec{h}}, \frac{\partial \vec{h}}{\partial \vec{v}}\right) = \boxed{\frac{\partial J}{\partial \vec{h}} \odot relu'(v)}$$

where:
$$relu'(v_i) = \begin{cases} 1, & \text{if } v_i > 0 \\ 0, & \text{otherwise} \end{cases}$$

EXAMPLE:

Given: $\vec{v} = \begin{bmatrix} -4 \\ 10 \\ 5 \end{bmatrix}_{3\times1}$

Assume: $\frac{\partial J}{\partial \vec{h}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{3\times1}$

then:
$$\frac{\partial J}{\partial \vec{v}} = \frac{\partial J}{\partial \vec{h}} \odot relu'(\vec{v}) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \odot \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \begin{matrix} \to relu'(-4) \\ \to relu'(10) \\ \to relu'(5) \end{matrix}$$

$$\boxed{\frac{\partial J}{\partial \vec{v}} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}_{3\times1}}$$

③ Regularization Layer

$\Rightarrow s = \lambda \|W\|_F^2$

Backpropagation:
→ have: $\frac{\partial J}{\partial s}$
→ goal: $\frac{\partial J}{\partial W}$

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

EXAMPLE:

Given:
$$W = \begin{bmatrix} 1 & 5 & 7 \\ 3 & 2 & 4 \end{bmatrix}_{2\times3} \qquad \lambda = 0.1$$

Assume: $\frac{\partial J}{\partial s} = 1$ (scalar)

then:
$$\frac{\partial J}{\partial W} = \frac{\partial J}{\partial s} \cdot \lambda W = 1 \cdot (0.1) \begin{bmatrix} 1 & 5 & 7 \\ 3 & 2 & 4 \end{bmatrix}$$

$$\boxed{\frac{\partial J}{\partial W} = \begin{bmatrix} 0.1 & 0.5 & 0.7 \\ 0.3 & 0.2 & 0.4 \end{bmatrix}_{2\times3}}$$

④ Sum Layer

$\Rightarrow K = s_1 + s_2$ (forward propagation)

Backpropagation:
→ have: $\frac{\partial J}{\partial K}$
→ goal: $\frac{\partial J}{\partial s_1}$, $\frac{\partial J}{\partial s_2}$

$$\frac{\partial J}{\partial s_1} = prod\left(\frac{\partial J}{\partial K}, \frac{\partial K}{\partial s_1}\right) = \boxed{\frac{\partial J}{\partial K}}$$
$$\frac{\partial J}{\partial s_2} = prod\left(\frac{\partial J}{\partial K}, \frac{\partial K}{\partial s_2}\right) = \boxed{\frac{\partial J}{\partial K}}$$

EXAMPLE:

Given: $s_1 = 50$, $s_2 = 5$

Assume: $\frac{\partial J}{\partial K} = 1$

then:
$$\frac{\partial J}{\partial s_1} = \frac{\partial J}{\partial K} \cdot \frac{\partial K}{\partial s_1} = 1 \cdot 1 \qquad \boxed{\frac{\partial J}{\partial s_1} = 1}$$
$$\frac{\partial J}{\partial s_2} = \frac{\partial J}{\partial K} \cdot \frac{\partial K}{\partial s_2} = 1 \cdot 1 \qquad \boxed{\frac{\partial J}{\partial s_2} = 1}$$

⑤ L_2 Loss Layer

$\Rightarrow L = \frac{1}{2} \|\vec{o} - \vec{y}\|_2^2$ (forward propagation)

Backpropagation:
→ have: $\frac{\partial J}{\partial L}$
→ goal: $\frac{\partial J}{\partial \vec{o}}$

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

EXAMPLE:

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

Assume: $\frac{\partial J}{\partial L} = 1$

then:
$$\frac{\partial J}{\partial \vec{o}} = \frac{\partial J}{\partial L} \cdot (\vec{o} - \vec{y}) = 1 \cdot \begin{bmatrix} 10 \\ 5 \end{bmatrix} - \begin{bmatrix} 11 \\ 5 \end{bmatrix}$$

$$\boxed{\frac{\partial J}{\partial \vec{o}} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}_{2\times1}}$$

⑥ Softmax (+ cross-entropy) Layer

$\Rightarrow \vec{o} = softmax(\vec{v})$
$L = cross\_entropy\_loss(\vec{o})$ (forward propagation)

Backpropagation:
→ have: $\frac{\partial J}{\partial L}$
→ goal: $\frac{\partial J}{\partial \vec{o}}$

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

EXAMPLE:

Given: $\vec{v} = \begin{bmatrix} 10 \\ 5 \end{bmatrix}_{2\times1}$, $\vec{y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{2\times1}$

Assume: $\frac{\partial J}{\partial L} = 1$

then:
→ have to calculate softmax of $\vec{v}$ to get $\vec{o}$ due to how I set up my tests

$$\vec{o} = softmax(\vec{v}) = softmax\left(\begin{bmatrix} 10 \\ 5 \end{bmatrix}\right)$$

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

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

$$\vec{o} = \begin{bmatrix} 0.993 \\ 0.007 \end{bmatrix}_{2\times1} \leftarrow \text{use for backprop}$$

$$\frac{\partial J}{\partial \vec{o}} = \frac{\partial J}{\partial L} \cdot (\vec{o} - \vec{y}) = 1 \cdot \left(\begin{bmatrix} 0.993 \\ 0.007 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$$

$$\boxed{\frac{\partial J}{\partial \vec{o}} = \begin{bmatrix} -0.007 \\ 0.007 \end{bmatrix}_{2\times1}}$$

Paige Rosynek

CS 3450

Lab 6

$\boxed{\text{Forward Propagation}}$

(unit tests)

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:

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

$$\vec{b} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}_{2 \times 1}$$

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

$$= \begin{bmatrix} (1)(2) + (5)(10) + (7)(5) \\ (3)(2) + (2)(10) + (4)(5) \end{bmatrix} + \begin{bmatrix} 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 87 \\ 46 \end{bmatrix} + \begin{bmatrix} 4 \\ 8 \end{bmatrix}$$

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

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:

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

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

$$= ReLU\left(\begin{bmatrix} -4 \\ 10 \\ 5 \end{bmatrix}\right) \Rightarrow \begin{aligned} ReLU(-4) &= 0 \\ ReLU(10) &= 10 \\ ReLU(5) &= 5 \end{aligned}$$

$$\boxed{\vec{h} = \begin{bmatrix} 0 \\ 10 \\ 5 \end{bmatrix}_{3 \times 1}}$$

3. Regularization Layer

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

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

where:

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

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

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:

$s, s_1, s_2 \in \mathbb{R}^1$

Example:

$s_1 = 50 \quad s_2 = 5$

$$s = 50 + 5$$

$$\boxed{s = 55}$$

5. L_2 Loss Layer (MSE)

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

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

where:

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

$\mathcal{L} \in \mathbb{R}$

Example

$$\vec{o} = \begin{bmatrix} 10 \\ 5 \end{bmatrix}_{2 \times 1}, \quad \vec{y} = \begin{bmatrix} 11 \\ 5 \end{bmatrix}_{2 \times 1}$$

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

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

$$= \frac{1}{2} (1 + 0)$$

$$\boxed{\mathcal{L} = 0.5}$$

6. Softmax Layer

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

where:

$\vec{o}, \vec{v} \in \mathbb{R}^{d \times 1}$

Note:

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

cross-entropy loss: $\mathcal{L}$

$$\mathcal{L} = CE(\vec{o}, \vec{y}) = -\frac{1}{N} \sum_{j=1}^{d} \left(y_j \cdot \log(o_j) \right)$$

Example

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

softmax:

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

where:

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

$$\vec{o} = softmax\left(\begin{bmatrix} 10 \\ 5 \end{bmatrix} \right)$$

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

$$= \frac{1}{1 + e^{-5}} = \underline{0.993}$$

$$softmax(5) = \frac{e^{5-10}}{e^{10-10} + e^{5-10}}$$

$$= \frac{e^{-5}}{1 + e^{-5}} = \underline{0.007}$$

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

check:

→ should sum to 1

$0.993 + 0.007 = 1$ ✓

cross-entropy loss:

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

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

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

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

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

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

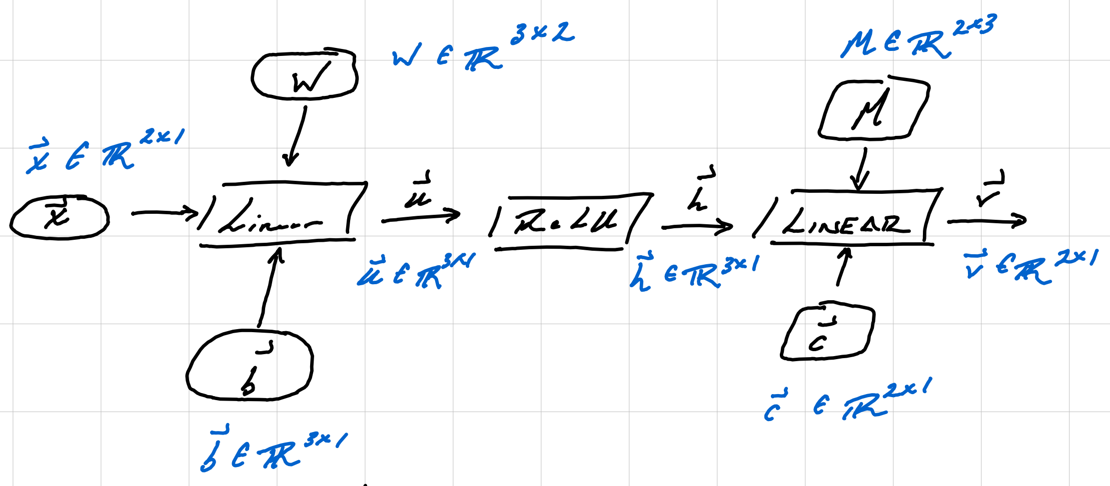
$$\boxed{\mathcal{L} = 0.003358}$$

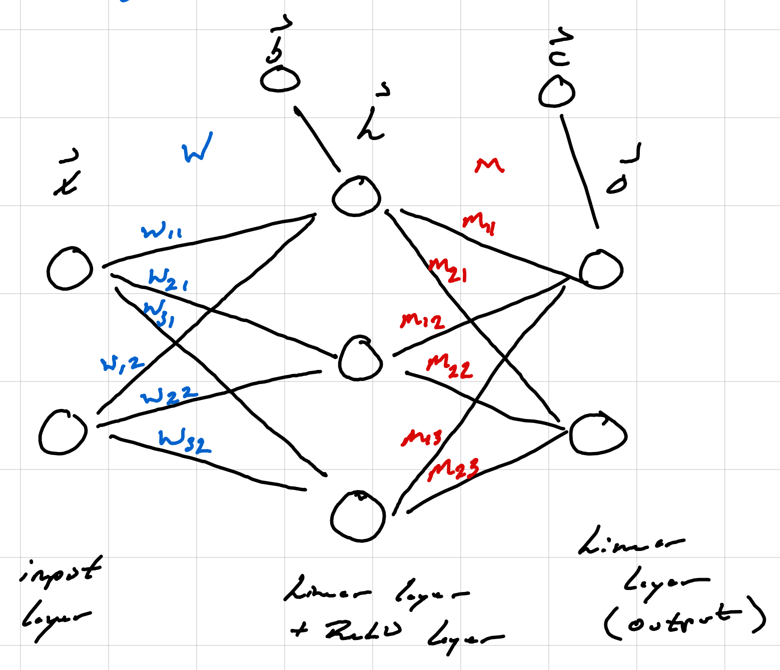
Note:

log → ln

FORWARD PROPAGATION
FULL NETWORK

network :





Given :

$$\vec{x} = \begin{bmatrix} 10 \\ 5 \end{bmatrix}_{2\times1}, \quad W = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 7 & 3 \end{bmatrix}_{3\times2}, \quad \vec{b} = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}_{3\times1}$$

$$\vec{c} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}_{2\times1}, \quad M = \begin{bmatrix} 4 & 1 & 2 \\ 3 & 6 & 1 \end{bmatrix}_{2\times3}$$

1. Linear Layer (hidden layer)

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

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

$$= \begin{bmatrix} (1)(10) + (2)(5) \\ (2)(10) + (5)(5) \\ (7)(10) + (3)(5) \end{bmatrix}_{3\times1} + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}_{3\times1}$$

$$= \begin{bmatrix} 20 \\ 45 \\ 85 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$$

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

ReLU :

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

$$= ReLU\left(\begin{bmatrix} 24 \\ 47 \\ 86 \end{bmatrix}\right)$$

$$ReLU(\overset{z_1}{24}) = 24$$

$$ReLU(\overset{z_2}{47}) = 47$$

$$ReLU(86) = 86$$

$$\boxed{\vec{h} = \begin{bmatrix} 24 \\ 47 \\ 86 \end{bmatrix}_{3\times1}}$$

2. Linear Layer (output-layer)

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

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

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

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

$$\boxed{\vec{o} = \begin{bmatrix} 320 \\ 441 \end{bmatrix}_{2\times1}}$$