Homework 02-1: due 2022/03/31 14:10 (30%)

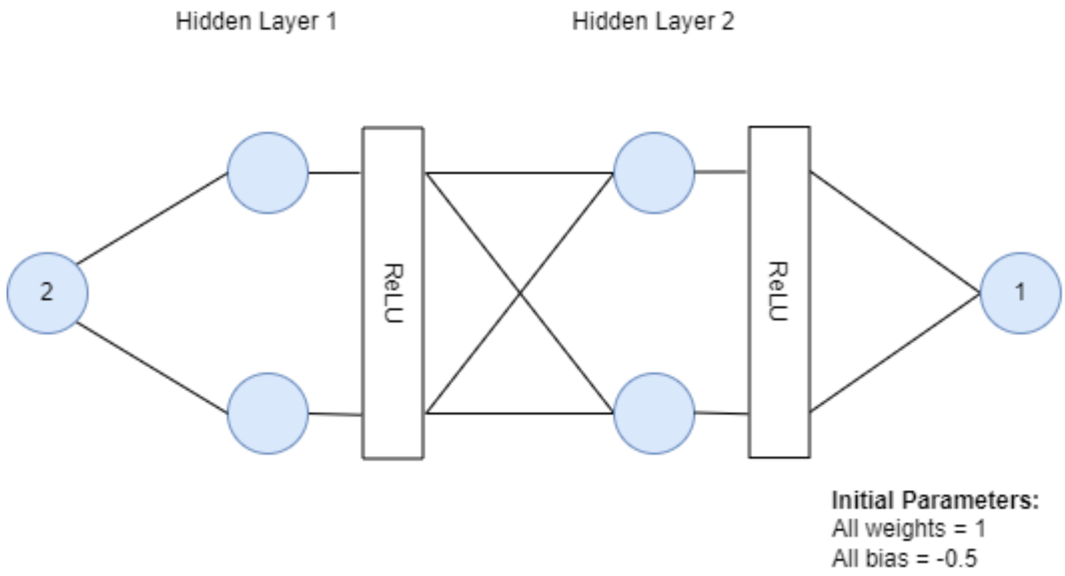
- In this part, you should calculate the forward pass and backpropagation manually and there is no need for any coding.

- Please scan your hand-writting calculation and save it as HW2-1.pdf

- By running the following script, you can check your answer and observe how to do the backpropagation in PyTorch.

- You can change the iterations in script to observe how will the loss change.

1. Please do the forward pass and backpropagation with a neural network as below, the input is 2 and the target is 1. Also, calculate the loss with half and the sum square error. Please update the parameters twice and use the learning rate 0.01.



In [8]:

```
import torch
import torch.nn as nn
import torch.optim as optim
from collections import OrderedDict
```

In [9]:

```
X = torch.tensor([2], dtype= torch.float32)
y = torch.tensor([1], dtype= torch.float32)
```

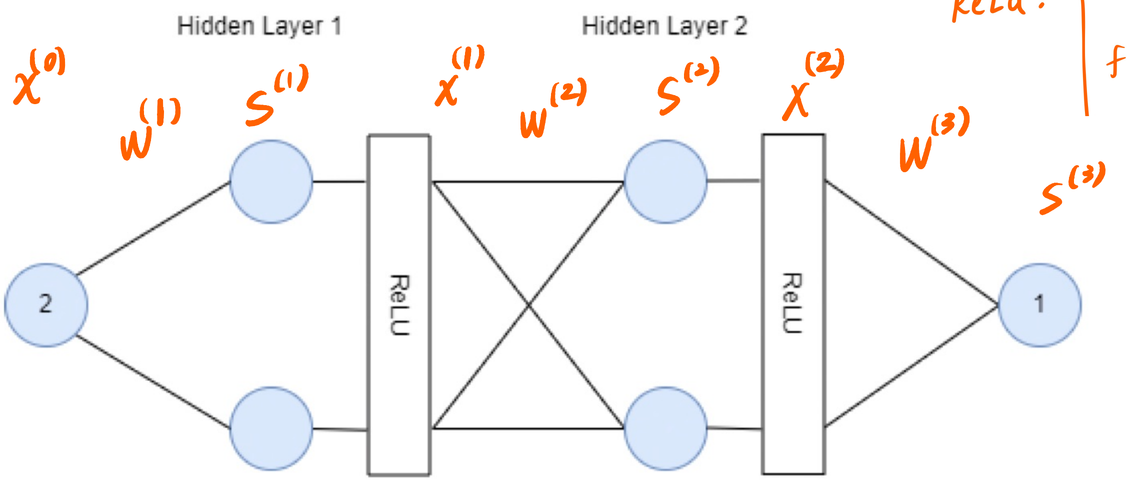
In [10]:

```
# Half of the sum square error
def loss(y, pred):
    return ((pred-y)**2).sum()/2
```

HW 2-1 calculation

1. Please do the forward pass and backpropagation with a neural network as below, the input is 2 and the target is 1. Also, calculate the loss with half and the sum square error. Please update the parameters twice and use the learning rate 0.01.

ReLU: $\begin{cases} f(x)=x, & x\geq 0 \\ f(x)=0, & x<0 \end{cases}$



Initial Parameters:
All weights = 1
All bias = -0.5

Forward Pass 1

$$\text{Input } \underset{\text{layer}}{x^{(0)}} = \begin{bmatrix} -0.5 \\ 2 \end{bmatrix} \text{(bias)}$$

Hidden Layer 1

$$S^{(1)} = (W^{(1)})^T \cdot X^{(0)} = \underset{\text{(weight)}}{\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}^T} \begin{bmatrix} -0.5 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix} \xrightarrow{\text{ReLU}} \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix}$$

$$\text{Input } x^{(1)} = \begin{bmatrix} -0.5 \\ 1.5 \\ 1.5 \end{bmatrix} \text{(bias)}$$

Hidden Layer 2

$$S^{(2)} = (W^{(2)})^T \cdot X^{(1)} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}^T \begin{bmatrix} -0.5 \\ 1.5 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 2.5 \\ 2.5 \end{bmatrix} \xrightarrow{\text{ReLU}} \begin{bmatrix} 2.5 \\ 2.5 \end{bmatrix}$$

$$\text{Input } X^{(2)} = \begin{bmatrix} -0.5 \\ 2.5 \\ 2.5 \end{bmatrix} \text{(bias)}$$

$$\text{output (prediction)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} -0.5 \\ 2.5 \\ 2.5 \end{bmatrix} = 4.5 = X^{(3)}$$

$$\text{Loss} = \frac{\Sigma(y-y^*)^2}{2} \Rightarrow \frac{(4.5-1)^2}{2} = 6.125$$

---

backpropagation I

Loss Function $(e) = \frac{1}{2}\Sigma(y-y^*)^2$

bias gradient

θ = activation function

θ' = derivative

$$\delta^{(3)} = \frac{\partial e}{\partial s^{(3)}} = (x^{(3)} - y^*)\frac{\partial x^{(3)}}{\partial s^{(3)}} = (x^{(3)} - y^*)\cdot\theta'(s^{(3)}) = 3.5$$

$$\delta^{(2)} = \theta'(s^{(2)}) \otimes \left[W^{(3)}\delta^{(3)}\right]_1^{d^{(2)}} = \begin{bmatrix}1\\1\end{bmatrix} \otimes \left(\begin{bmatrix}1\\1\end{bmatrix}\cdot 3.5\right) = \begin{bmatrix}3.5\\3.5\end{bmatrix}$$

$$\delta^{(1)} = \theta'(s^{(1)}) \otimes \left[W^{(2)}\delta^{(2)}\right]_1^{d(1)} = \begin{bmatrix}1\\1\end{bmatrix} \otimes \left(\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\cdot\begin{bmatrix}3.5\\3.5\end{bmatrix}\right) = \begin{bmatrix}1\\1\end{bmatrix} \otimes \begin{bmatrix}7\\7\end{bmatrix} = \begin{bmatrix}7\\7\end{bmatrix}$$

δ : sensitivity vector

→ gradient of the error (e) with respect to the input signal $s^{(l)}$

$$\frac{\partial e}{\partial w^{(l)}} = \underbrace{x^{(l-1)}}_{\text{include bias}}\cdot(\delta^{(l)})^T$$

weight gradient

$$\frac{\partial e}{\partial w^{(1)}} = x^{(0)}\cdot(\delta^{(1)})^T = \begin{bmatrix}-0.5\\2\end{bmatrix}\begin{bmatrix}7\\7\end{bmatrix}^T = \begin{bmatrix}-3.5 & -3.5\\14 & 14\end{bmatrix}$$

$$\frac{\partial e}{\partial w^{(2)}} = x^{(1)}\cdot(\delta^{(2)})^T = \begin{bmatrix}-0.5\\1.5\\1.5\end{bmatrix}\begin{bmatrix}3.5 & 3.5\end{bmatrix} = \begin{bmatrix}-1.75 & -1.75\\5.25 & 5.25\\5.25 & 5.25\end{bmatrix}$$

$$\frac{\partial e}{\partial w^{(3)}} = x^{(2)}\cdot(\delta^{(3)})^T = \begin{bmatrix}-0.5\\2.5\\2.5\end{bmatrix}\cdot 3.5 = \begin{bmatrix}-1.75\\8.75\\8.75\end{bmatrix}$$

Updated Weight

η (learning rate) $= 0.01$

= updated weight

$$W^{(\ell)}_{(new)} = W^{(\ell)} - \eta \frac{\partial e}{\partial W^{(\ell)}}$$

iteration number

$$W^{(1)}(1) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - 0.01 \begin{bmatrix} -3.5 & -3.5 \\ 14 & 14 \end{bmatrix} = \begin{bmatrix} 1.035 & 1.035 \\ 0.86 & 0.86 \end{bmatrix}$$

$$W^{(2)}(1) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} - 0.01 \begin{bmatrix} -1.75 & -1.75 \\ 5.75 & 5.75 \\ 5.75 & 5.75 \end{bmatrix} = \begin{bmatrix} 1.0175 & 1.0175 \\ 0.9475 & 0.9475 \\ 0.9475 & 0.9475 \end{bmatrix}$$

$$W^{(3)}(1) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 0.01 \begin{bmatrix} -1.75 \\ 8.75 \\ 8.75 \end{bmatrix} = \begin{bmatrix} 1.0175 \\ 0.9125 \\ 0.9125 \end{bmatrix}$$

Updated bias

$$b^{(\ell)}_{(new)} = b^{(\ell)} - \eta \cdot (\text{bias gradient}^{(\ell)})$$

$$b^{(1)}(1) = -0.5 - 0.01 \cdot 7 = -0.57$$

$$b^{(2)}(1) = -0.5 - 0.01 \cdot 3.5 = -0.535$$

$$b^{(3)}(1) = -0.5 - 0.01 \cdot 3.5 = -0.535$$

Forward Pass 2

Input $x^{(0)}(1) = \begin{bmatrix} -0.57 \\ 2 \end{bmatrix}$ (bias)

updated weight matrix

Hidden Layer 1: $s^{(1)}(1) = \begin{bmatrix} 1 & 0.86 \\ 1 & 0.86 \end{bmatrix} \begin{bmatrix} -0.57 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.15 \\ 1.15 \end{bmatrix} \xrightarrow{\text{ReLU}} \begin{bmatrix} 1.15 \\ 1.15 \end{bmatrix}$

bias

Input $x^{(1)}(1) = \begin{bmatrix} -0.535 \\ 1.15 \\ 1.15 \end{bmatrix}$

Hidden Layer 2: $s^{(2)}(1) = \begin{bmatrix} 1 & 0.9475 & 0.9475 \\ 1 & 0.9475 & 0.9475 \end{bmatrix} \begin{bmatrix} -0.535 \\ 1.15 \\ 1.15 \end{bmatrix} = \begin{bmatrix} 1.64425 \\ 1.64425 \end{bmatrix}$

$$\text{Input } x^{(2)}_{(1)} = \begin{bmatrix} -0.535 \\ 1.64425 \\ 1.64425 \end{bmatrix}$$

$$\text{output (prediction)} = \begin{bmatrix} 1 & 0.9125 & 0.9125 \end{bmatrix} \begin{bmatrix} -0.535 \\ 1.64425 \\ 1.64425 \end{bmatrix} = 2.46575 \doteq 2.4658$$

---

backpropagation 2

Bias Gradient

$$\delta^{(3)}_{(1)} = (2.4658 - 1) \cdot 1 = 1.4658$$

$$\delta^{(2)}_{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \left(\begin{bmatrix} 0.9125 \\ 0.9125 \end{bmatrix} \cdot 1.4658 \right) = \begin{bmatrix} 1.3375425 \\ 1.3375425 \end{bmatrix} \doteq \begin{bmatrix} 1.3375 \\ 1.3375 \end{bmatrix}$$

$$\delta^{(1)}_{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \left(\begin{bmatrix} 0.9475 & 0.9475 \\ 0.9475 & 0.9475 \end{bmatrix} \begin{bmatrix} 1.3375 \\ 1.3375 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 2.5345625 \\ 2.5345625 \end{bmatrix}$$

Weight Gradient

$$\frac{\partial e}{\partial w^{(1)}_{(1)}} = x^{(0)}_{(1)} \cdot \left(\delta^{(1)}_{(1)} \right)^T = \begin{bmatrix} -0.57 \\ 2 \end{bmatrix} \begin{bmatrix} 2.5346 & 2.5346 \end{bmatrix} = \begin{bmatrix} -1.4448 & -1.4448 \\ 5.0691 & 5.0691 \end{bmatrix}$$

$$\frac{\partial e}{\partial w^{(2)}_{(1)}} = x^{(1)}_{(1)} \cdot \left(\delta^{(2)}_{(1)} \right)^T = \begin{bmatrix} -0.535 \\ 1.15 \\ 1.15 \end{bmatrix} \begin{bmatrix} 1.3375 & 1.3375 \end{bmatrix}$$

$$= \begin{bmatrix} -0.7156 & -0.7156 \\ 1.5381 & 1.5381 \\ 1.5381 & 1.5381 \end{bmatrix}$$

$$\frac{\partial e}{\partial w^{(3)}_{(1)}} = x^{(2)}_{(1)} \cdot \left(\delta^{(3)}_{(1)} \right)^T = \begin{bmatrix} -0.535 \\ 1.64425 \\ 1.64425 \end{bmatrix} \cdot 1.4658 = \begin{bmatrix} -0.7842 \\ 2.4101 \\ 2.4101 \end{bmatrix}$$

Weight gradient

Update Weight $\eta = 0.01$

$$W^{(1)}(2) = \begin{bmatrix} 1 & 1 \\ 0.86 & 0.86 \end{bmatrix} - 0.01 \begin{bmatrix} -1.4448 & -1.4448 \\ 5.0691 & 5.0691 \end{bmatrix} = \begin{bmatrix} 1.0144 & 1.0144 \\ 0.8093 & 0.8093 \end{bmatrix}$$

$$W^{(2)}(2) = \begin{bmatrix} 1 & 1 \\ 0.9475 & 0.9475 \\ 0.9475 & 0.9475 \end{bmatrix} - 0.01 \begin{bmatrix} -0.7156 & -0.7156 \\ 1.5381 & 1.5381 \\ 1.5381 & 1.5381 \end{bmatrix} = \begin{bmatrix} 1.0072 & 1.0072 \\ 0.9321 & 0.9321 \\ 0.9321 & 0.9321 \end{bmatrix}$$

$$W^{(3)}(2) = \begin{bmatrix} 1 \\ 0.9125 \\ 0.9125 \end{bmatrix} - 0.01 \begin{bmatrix} -0.7482 \\ 2.4101 \\ 2.4101 \end{bmatrix} = \begin{bmatrix} 1.0074 \\ 0.8884 \\ 0.8884 \end{bmatrix}$$

Update Bias $\eta = 0.01$

$$b^{(1)}(2) = -0.57 - 0.01 \cdot 2.5346 = -0.595346 \doteq -0.5953$$

$$b^{(2)}(2) = -0.535 - 0.01 \cdot 1.3375 = -0.548375 \doteq -0.5484$$

$$b^{(3)}(2) = -0.535 - 0.01 \cdot 1.4658 = -0.549658 \doteq -0.5497$$

Forward pass 3 (更新第2次)

Input $x^{(0)}(2) = \begin{bmatrix} -0.5953 \\ 2 \end{bmatrix}$

Hidden Layer 1 : $s^{(1)}(2) = \begin{bmatrix} 1 & 0.8093 \\ 1 & 0.8093 \end{bmatrix} \begin{bmatrix} -0.5953 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.0233 \\ 1.0233 \end{bmatrix} \xrightarrow{\text{ReLU}} \begin{bmatrix} 1.0233 \\ 1.0233 \end{bmatrix}$

Input $x^{(1)}(2) = \begin{bmatrix} -0.5484 \\ 1.0233 \\ 1.0233 \end{bmatrix}$

Hidden Layer 2 : $s^{(2)}(2) = \begin{bmatrix} 1 & 0.9321 & 0.9321 \\ 1 & 0.9321 & 0.9321 \end{bmatrix} \begin{bmatrix} -0.5484 \\ 1.0233 \\ 1.0233 \end{bmatrix}$

$$= \begin{bmatrix} 1.35923586 \\ 1.35923586 \end{bmatrix} \xrightarrow{\text{ReLU}} \begin{bmatrix} 1.3592 \\ 1.3592 \end{bmatrix}$$

Input $x^{(2)}(1) = \begin{bmatrix} -0.5497 \\ 1.3592 \\ 1.3592 \end{bmatrix}$

$$\text{output (prediction)} = \begin{bmatrix} 1 & 0.8884 & 0.8884 \end{bmatrix} \begin{bmatrix} -0.5497 \\ 1.3592 \\ 1.3592 \end{bmatrix}$$

$$= 1.86532656$$

$$\doteq 1.8654$$

$$\text{Loss} = \frac{(1.8654-1)^2}{2} = 0.37445858 \doteq 0.3745$$