universität freiburg

Assignment 07

Neural Networks – Part 2

Machine Learning WS23/24

Assignment 07
Solution

Backpropagation

Backpropagation

Task 1.1: Advantage of Backpropagation

What is the advantage of backpropagation vs. computing the gradients in isolation for every parameter?

- We can re-use intermediate results
- In particular: We can re-use gradient calculations from the previous layer

Backpropagation

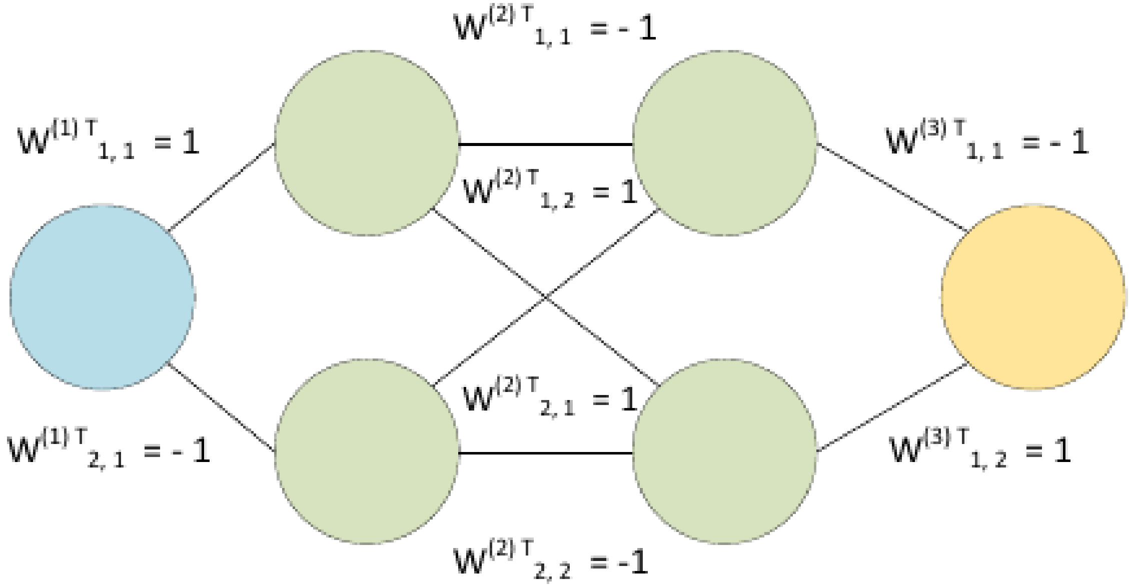
Task 1.2: Perform Backpropagation

Given Network:

$$
\begin{aligned}
h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right), & a^{(1)} &= W^{(1)} \cdot x \\
h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right), & a^{(2)} &= W^{(2)} \cdot h^{(1)} \\
h^{(3)} &= W^{(3)} \cdot h^{(2)}
\end{aligned}
$$

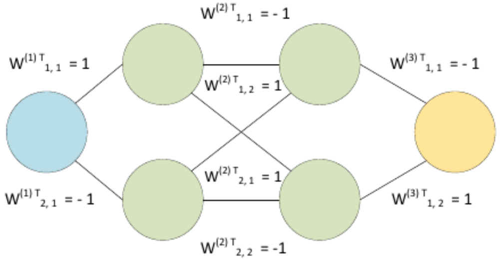
Given Weights:

$$
\begin{aligned}
W^{(1)} &= \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
W^{(2)} &= \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \\
W^{(3)} &= \begin{pmatrix} -1 & 1 \end{pmatrix}
\end{aligned}
$$



Backpropagation

Task 1.2: Perform Backpropagation



Given Network:

$$
\begin{aligned}
h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right), & a^{(1)} &= W^{(1)} \cdot x \\
h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right), & a^{(2)} &= W^{(2)} \cdot h^{(1)} \\
h^{(3)} &= W^{(3)} \cdot h^{(2)}
\end{aligned}
$$

Given Weights:

$$
\begin{aligned}
W^{(1)} &= \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
W^{(2)} &= \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \\
W^{(3)} &= \begin{pmatrix} -1 & 1 \end{pmatrix}
\end{aligned}
$$

1. Compute Forward Pass (x = 3, y = 6)

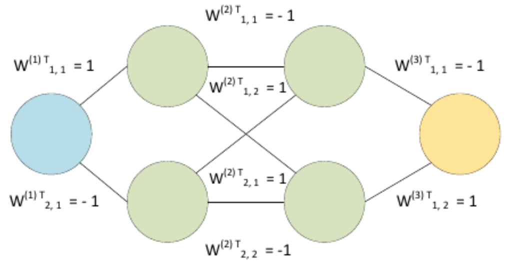
$$
\begin{aligned}
a^{(1)} &= W^{(1)} \cdot x = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \cdot 3 = \begin{pmatrix} 3 \\ -3 \end{pmatrix} \\
h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right) = \mathrm{ReLU}\left(\begin{pmatrix} 3 \\ -3 \end{pmatrix}\right) = \begin{pmatrix} 3 \\ 0 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
a^{(2)} &= W^{(2)} \cdot h^{(1)} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 0 \end{pmatrix} = \begin{pmatrix} -3 \\ 3 \end{pmatrix} \\
h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right) = \mathrm{ReLU}\left(\begin{pmatrix} -3 \\ 3 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 3 \end{pmatrix}
\end{aligned}
$$

$$
h^{(3)} = W^{(3)} \cdot x = \begin{pmatrix} -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 3 \end{pmatrix} = 3
$$

Backpropagation

Task 1.2: Perform Backpropagation



Given Network:

$$
\begin{aligned}
h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right), & a^{(1)} &= W^{(1)} \cdot x \\
h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right), & a^{(2)} &= W^{(2)} \cdot h^{(1)} \\
h^{(3)} &= W^{(3)} \cdot h^{(2)}
\end{aligned}
$$

2: Compute Loss (x = 3, y = 6)

$$
\begin{aligned}
J(w) &= (y - \hat{y})^2 \\
&= (6-3)^2 \\
&= 9
\end{aligned}
$$

Given Weights:

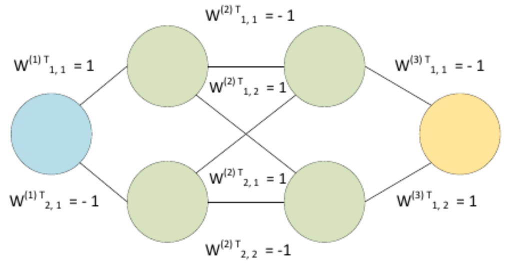
$$
\begin{aligned}
W^{(1)} &= \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
W^{(2)} &= \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \\
W^{(3)} &= \begin{pmatrix} -1 & 1 \end{pmatrix}
\end{aligned}
$$

From prev. Step:

$$h^{(3)} = \hat{y} = 3$$

Backpropagation

Task 1.2: Perform Backpropagation



Given Network:

$$\begin{aligned} h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right), & a^{(1)} &= W^{(1)} \cdot x \\ h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right), & a^{(2)} &= W^{(2)} \cdot h^{(1)} \\ h^{(3)} &= W^{(3)} \cdot h^{(2)} \end{aligned}$$

Given Weights:

$$W^{(1)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$W^{(2)} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$W^{(3)} = \begin{pmatrix} -1 & 1 \end{pmatrix}$$

3: Backpropagation

What do we actually need to compute?

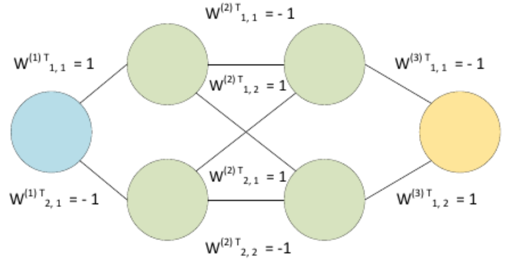
- The gradients for each weight wrt. Loss:

$$\frac{\partial J(w)}{\partial w^{(i)}_{j,k}} = ? \qquad \text{for all } i, j, k$$

As the name implies: Start with the loss and propagate results backwards.

Backpropagation

Task 1.2: Perform Backpropagation



Given Network:

$$
\begin{aligned}
h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right), & a^{(1)} &= W^{(1)} \cdot x \\
h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right), & a^{(2)} &= W^{(2)} \cdot h^{(1)} \\
h^{(3)} &= W^{(3)} \cdot h^{(2)}
\end{aligned}
$$

Given Weights:

$$
\begin{aligned}
W^{(1)} &= \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
W^{(2)} &= \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \\
W^{(3)} &= \begin{pmatrix} -1 & 1 \end{pmatrix}
\end{aligned}
$$

From prev. Steps:

$$
h^{(3)} = \hat{y} = 3
$$

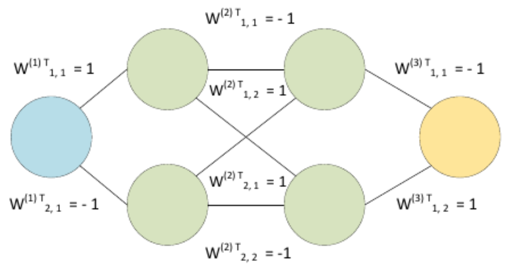
3: Backpropagation – Part 1: Loss

$$
J(w) = \left(y - \hat{y}\right)^2
$$

$$
\begin{aligned}
\frac{\partial J(w)}{\partial h^{(3)}} &= -2 \cdot \left(y - h^{(3)}\right) \\
&= -2 \cdot (6 - 3) \\
&= -6
\end{aligned}
$$

Backpropagation

Task 1.2: Perform Backpropagation



Given Network:

$$
\begin{aligned}
h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right), & a^{(1)} &= W^{(1)} \cdot x \\
h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right), & a^{(2)} &= W^{(2)} \cdot h^{(1)} \\
h^{(3)} &= W^{(3)} \cdot h^{(2)}
\end{aligned}
$$

Given Weights:

$$
W^{(1)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}
$$

$$
W^{(2)} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}
$$

$$
W^{(3)} = \begin{pmatrix} -1 & 1 \end{pmatrix}
$$

From prev. Steps:

$$
\frac{\partial J(w)}{\partial h^{(3)}} = -6
$$

$$
h^{(2)} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}
$$

3: Backpropagation – Part 2: Layer 3 Weights

$$
\begin{aligned}
\frac{\partial J(w)}{\partial w^{(3)}_{1,1}} &= \frac{\partial J(w)}{\partial h^{(3)}} \cdot \frac{\partial h^{(3)}}{\partial w^{(3)}_{1,1}} \\
\frac{\partial J(w)}{\partial w^{(3)}_{1,2}} &= \frac{\partial J(w)}{\partial h^{(3)}} \cdot \frac{\partial h^{(3)}}{\partial w^{(3)}_{1,2}}
\end{aligned}
$$

apply chain rule

$$
\begin{aligned}
\frac{\partial h^{(3)}}{\partial w^{(3)}_{1,1}} &= \frac{\partial}{\partial w^{(3)}_{1,1}} \left(W^{(3)} h^{(2)}\right) = h^{(2)}_1 \\
\frac{\partial h^{(3)}}{\partial w^{(3)}_{1,2}} &= \frac{\partial}{\partial w^{(3)}_{1,2}} \left(W^{(3)} h^{(2)}\right) = h^{(2)}_2
\end{aligned}
$$

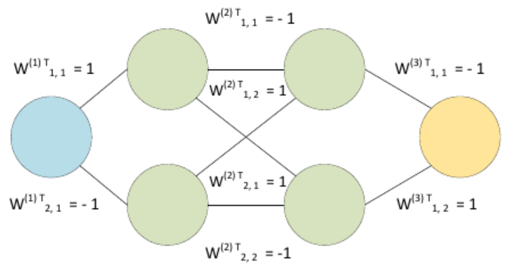
compute new gradients

$$
\begin{aligned}
\frac{\partial J(w)}{\partial w^{(3)}_{1,1}} &= -6 \cdot 0 = 0 \\
\frac{\partial J(w)}{\partial w^{(3)}_{1,2}} &= -6 \cdot 3 = -18
\end{aligned}
$$

plug in

Backpropagation

Task 1.2: Perform Backpropagation



Given Network:

$$
\begin{aligned}
h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right), & a^{(1)} &= W^{(1)} \cdot x \\
h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right), & a^{(2)} &= W^{(2)} \cdot h^{(1)} \\
h^{(3)} &= W^{(3)} \cdot h^{(2)}
\end{aligned}
$$

Given Weights:

$$
W^{(1)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}
$$

$$
W^{(2)} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}
$$

$$
W^{(3)} = \begin{pmatrix} -1 & 1 \end{pmatrix}
$$

From prev. Steps:

$$
\frac{\partial J(w)}{\partial h^{(3)}} = -6
$$

3: Backpropagation – Part 3: Layer 2 Activation Inputs

$$
\begin{aligned}
\frac{\partial J(w)}{\partial a_1^{(2)}} &= \frac{\partial J(w)}{\partial h^{(3)}} \cdot \frac{\partial h^{(3)}}{\partial h_1^{(2)}} \cdot \frac{\partial h_1^{(2)}}{\partial a_1^{(2)}} \\
\frac{\partial J(w)}{\partial a_2^{(2)}} &= \frac{\partial J(w)}{\partial h^{(3)}} \cdot \frac{\partial h^{(3)}}{\partial h_2^{(2)}} \cdot \frac{\partial h_2^{(2)}}{\partial a_2^{(2)}}
\end{aligned}
$$

apply chain rule

$$
\begin{aligned}
\frac{\partial h^{(3)}}{\partial h_1^{(2)}} &= w_{1,1}^{(3)} & \frac{\partial h_1^{(2)}}{\partial a_1^{(2)}} &= 1_{a_1^{(2)} > 0} \\
\frac{\partial h^{(3)}}{\partial h_2^{(2)}} &= w_{1,2}^{(3)} & \frac{\partial h_2^{(2)}}{\partial a_2^{(2)}} &= 1_{a_2^{(2)} > 0}
\end{aligned}
$$

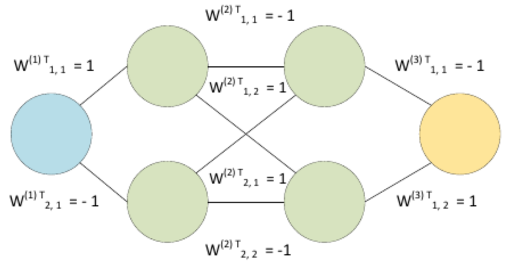
compute new gradients

$$
\begin{aligned}
\frac{\partial J(w)}{\partial a_1^{(2)}} &= -6 \cdot (-1) \cdot 0 = 0 \\
\frac{\partial J(w)}{\partial a_2^{(2)}} &= -6 \cdot 1 \cdot 1 = -6
\end{aligned}
$$

plug in

Backpropagation

Task 1.2: Perform Backpropagation



Given Network:

$$
\begin{aligned}
h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right), & a^{(1)} &= W^{(1)} \cdot x \\
h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right), & a^{(2)} &= W^{(2)} \cdot h^{(1)} \\
h^{(3)} &= W^{(3)} \cdot h^{(2)}
\end{aligned}
$$

Given Weights:

$$
\begin{aligned}
W^{(1)} &= \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
W^{(2)} &= \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \\
W^{(3)} &= \begin{pmatrix} -1 & 1 \end{pmatrix}
\end{aligned}
$$

From prev. Steps:

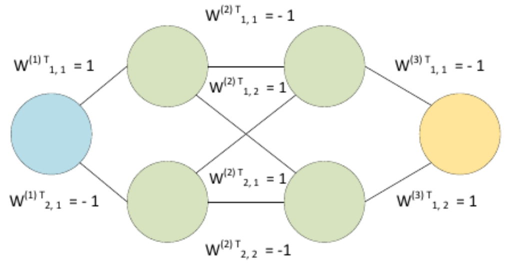
$$
\begin{aligned}
\frac{\partial J(w)}{\partial a_1^{(2)}} &= 0 \\
\frac{\partial J(w)}{\partial a_2^{(2)}} &= -6 \\
h^{(1)} &= \begin{pmatrix} 3 \\ 0 \end{pmatrix}
\end{aligned}
$$

3: Backpropagation – Part 4: Layer 2 Weights

$$
\begin{aligned}
\frac{\partial J(w)}{\partial w_{1,1}^{(2)}} &= \frac{\partial J(w)}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial w_{1,1}^{(2)}} = \frac{\partial J(w)}{\partial a_1^{(2)}} \cdot h_1^{(1)} = 0 \cdot 3 = 0 \\
\frac{\partial J(w)}{\partial w_{1,2}^{(2)}} &= \frac{\partial J(w)}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial w_{1,2}^{(2)}} = \frac{\partial J(w)}{\partial a_1^{(2)}} \cdot h_2^{(1)} = 0 \cdot 0 = 0 \\
\frac{\partial J(w)}{\partial w_{2,1}^{(2)}} &= \frac{\partial J(w)}{\partial a_2^{(2)}} \cdot \frac{\partial a_2^{(2)}}{\partial w_{2,1}^{(2)}} = \frac{\partial J(w)}{\partial a_2^{(2)}} \cdot h_1^{(1)} = -6 \cdot 3 = -18 \\
\frac{\partial J(w)}{\partial w_{2,2}^{(2)}} &= \frac{\partial J(w)}{\partial a_2^{(2)}} \cdot \frac{\partial a_2^{(2)}}{\partial w_{2,2}^{(2)}} = \frac{\partial J(w)}{\partial a_2^{(2)}} \cdot h_2^{(1)} = -6 \cdot 0 = 0
\end{aligned}
$$

Backpropagation

Task 1.2: Perform Backpropagation



Given Network:

$$
\begin{aligned}
h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right), & a^{(1)} &= W^{(1)} \cdot x \\
h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right), & a^{(2)} &= W^{(2)} \cdot h^{(1)} \\
h^{(3)} &= W^{(3)} \cdot h^{(2)}
\end{aligned}
$$

Given Weights:

$$
W^{(1)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}
$$

$$
W^{(2)} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}
$$

$$
W^{(3)} = \begin{pmatrix} -1 & 1 \end{pmatrix}
$$

From prev. Steps:

$$
\frac{\partial J(w)}{\partial a_1^{(2)}} = 0
$$

$$
\frac{\partial J(w)}{\partial a_2^{(2)}} = -6
$$

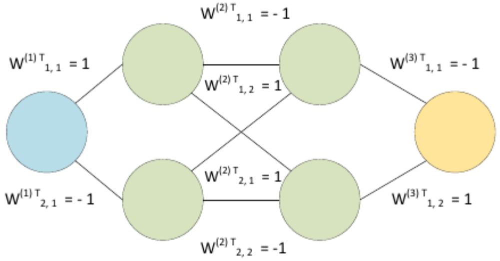
$$
a^{(1)} = \begin{pmatrix} 3 \\ -3 \end{pmatrix}
$$

3: Backpropagation – Part 5: Layer 1 Activation Inputs (1)

$$
\begin{aligned}
\frac{\partial J(w)}{\partial a_1^{(1)}} &= \frac{\partial J(w)}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial a_1^{(1)}} + \frac{\partial J(w)}{\partial a_2^{(2)}} \cdot \frac{\partial a_2^{(2)}}{\partial a_1^{(1)}} \\
&= \frac{\partial J(w)}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial h_1^{(1)}} \cdot \frac{\partial h_1^{(1)}}{\partial a_1^{(1)}} + \frac{\partial J(w)}{\partial a_2^{(2)}} \cdot \frac{\partial a_2^{(2)}}{\partial h_1^{(1)}} \cdot \frac{\partial h_1^{(1)}}{\partial a_1^{(1)}} \\
&= \frac{\partial J(w)}{\partial a_1^{(2)}} \cdot w_{1,1}^{(2)} \cdot 1_{a_1^{(1)} > 0} + \frac{\partial J(w)}{\partial a_2^{(2)}} \cdot w_{2,1}^{(2)} \cdot 1_{a_1^{(1)} > 0} \\
&= 0 \cdot (-1) \cdot 1 + (-6) \cdot 1 \cdot 1 \\
&= -6
\end{aligned}
$$

Backpropagation

Task 1.2: Perform Backpropagation



Given Network:

$$
\begin{aligned}
h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right), & a^{(1)} &= W^{(1)} \cdot x \\
h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right), & a^{(2)} &= W^{(2)} \cdot h^{(1)} \\
h^{(3)} &= W^{(3)} \cdot h^{(2)}
\end{aligned}
$$

Given Weights:

$$
\begin{aligned}
W^{(1)} &= \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
W^{(2)} &= \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \\
W^{(3)} &= \begin{pmatrix} -1 & 1 \end{pmatrix}
\end{aligned}
$$

From prev. Steps:

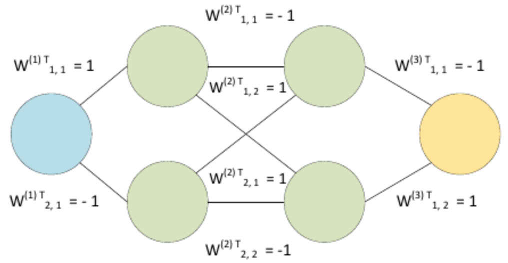
$$
\begin{aligned}
\frac{\partial J(w)}{\partial a_1^{(2)}} &= 0 \\
\frac{\partial J(w)}{\partial a_2^{(2)}} &= -6 \\
a^{(1)} &= \begin{pmatrix} 3 \\ -3 \end{pmatrix}
\end{aligned}
$$

3: Backpropagation – Part 5: Layer 1 Activation Inputs (2)

$$
\begin{aligned}
\frac{\partial J(w)}{\partial a_2^{(1)}} &= \frac{\partial J(w)}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial a_2^{(1)}} + \frac{\partial J(w)}{\partial a_2^{(2)}} \cdot \frac{\partial a_2^{(2)}}{\partial a_2^{(1)}} \\
&= \frac{\partial J(w)}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial h_2^{(1)}} \cdot \frac{\partial h_2^{(1)}}{\partial a_2^{(1)}} + \frac{\partial J(w)}{\partial a_2^{(2)}} \cdot \frac{\partial a_2^{(2)}}{\partial h_2^{(1)}} \cdot \frac{\partial h_2^{(1)}}{\partial a_2^{(1)}} \\
&= \frac{\partial J(w)}{\partial a_1^{(2)}} \cdot w_{1,2}^{(2)} \cdot 1_{a_2^{(1)} > 0} + \frac{\partial J(w)}{\partial a_2^{(2)}} \cdot w_{2,2}^{(2)} \cdot 1_{a_2^{(1)} > 0} \\
&= 0 \cdot 1 \cdot 0 + (-6) \cdot (-1) \cdot 0 \\
&= 0
\end{aligned}
$$

Backpropagation

Task 1.2: Perform Backpropagation



Given Network:

$$
\begin{aligned}
h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right), & a^{(1)} &= W^{(1)} \cdot x \\
h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right), & a^{(2)} &= W^{(2)} \cdot h^{(1)} \\
h^{(3)} &= W^{(3)} \cdot h^{(2)}
\end{aligned}
$$

Given Weights:

$$
\begin{aligned}
W^{(1)} &= \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
W^{(2)} &= \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \\
W^{(3)} &= \begin{pmatrix} -1 & 1 \end{pmatrix}
\end{aligned}
$$

From prev. Steps:

$$
\begin{aligned}
\frac{\partial J(w)}{\partial a_1^{(1)}} &= -6 \\
\frac{\partial J(w)}{\partial a_2^{(1)}} &= 0 \\
x &= 3
\end{aligned}
$$

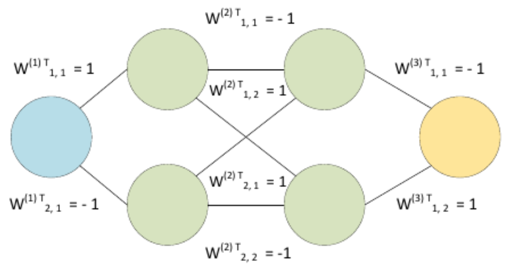
3: Backpropagation – Part 6: Layer 1 Weights

$$
\frac{\partial J(w)}{\partial w_{1,1}^{(1)}} = \frac{\partial J(w)}{\partial a_1^{(1)}} \cdot \frac{\partial a_1^{(1)}}{\partial w_{1,1}^{(1)}} = \frac{\partial J(w)}{\partial a_1^{(1)}} \cdot x = -6 \cdot 3 = -18
$$

$$
\frac{\partial J(w)}{\partial w_{1,2}^{(1)}} = \frac{\partial J(w)}{\partial a_2^{(1)}} \cdot \frac{\partial a_2^{(1)}}{\partial w_{1,2}^{(1)}} = \frac{\partial J(w)}{\partial a_2^{(1)}} \cdot x = 0 \cdot 3 = 0
$$

Backpropagation

Task 1.2: Perform Backpropagation



Given Network:

$$
\begin{aligned}
h^{(1)} &= \mathrm{ReLU}\left(a^{(1)}\right), & a^{(1)} &= W^{(1)} \cdot x \\
h^{(2)} &= \mathrm{ReLU}\left(a^{(2)}\right), & a^{(2)} &= W^{(2)} \cdot h^{(1)} \\
h^{(3)} &= W^{(3)} \cdot h^{(2)}
\end{aligned}
$$

Given Weights:

$$
\begin{aligned}
W^{(1)} &= \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
W^{(2)} &= \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \\
W^{(3)} &= \begin{pmatrix} -1 & 1 \end{pmatrix}
\end{aligned}
$$

From prev. Steps:

$$
\begin{aligned}
\frac{\partial J(w)}{\partial w^{(3)}_{1,2}} &= -18 \\
\frac{\partial J(w)}{\partial w^{(2)}_{2,1}} &= -18 \\
\frac{\partial J(w)}{\partial w^{(1)}_{1,1}} &= -18
\end{aligned}
$$

4: Update Weights

- General:

$$
w^{(i)}_{j,k} \leftarrow w^{(i)}_{j,k} - \eta \cdot \frac{\partial J(w)}{\partial w^{(i)}_{j,k}}
$$

- Only $w^{(3)}_{1,2}$, $w^{(2)}_{2,1}$, and $w^{(1)}_{1,1}$ are nonzero, therefore

$$
\begin{aligned}
w^{(3)}_{1,2} &\leftarrow 1 - 0.1 \cdot (-18) = 1.18 \\
w^{(2)}_{2,1} &\leftarrow 1 - 0.1 \cdot (-18) = 1.18 \\
w^{(1)}_{1,1} &\leftarrow 1 - 0.1 \cdot (-18) = 1.18
\end{aligned}
$$

- All other weights don’t change

Convolutional Neural Networks

Convolutional Neural Networks
Task 2.1: CNNs vs. Feed-Forward Networks

What is the main structural difference between CNNs and Feed Forward Neural Networks?

- Feed-forward neural networks:
  - Units are fully connected
  - Every unit/neuron on layer *l* is connected to all previous units on layer *l-1*
- Convolutional neural networks:
  - Sliding kernels: Local connectivity only
  - Pattern matching: Kernel is applied over the input, pattern is valued the same at every location

Convolutional Neural Networks

Task 2.2: Output Dimensions

Given a CNN with

- Input image of size 4 x 4 (grayscale)
- 2 kernels with size 2 x 2
- No padding
- Stride 1 or 2

What are the dimensions for stride 1?

- 2 x 3 x 3

What are the dimensions for stride 2?

- 2 x 2 x 2

Convolutional Neural Networks
Task 2.3: Computing Convolution

Given image *I* and kernel *K* with

$$I = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 5 & 3 \end{pmatrix}$$

$$K = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Compute convolution $O = I * K$

$$O = \begin{pmatrix} o_{1,1} & o_{1,2} \\ o_{2,1} & o_{2,2} \end{pmatrix}$$

$$o_{1,1}: \quad \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \rightarrow o_{1,1} = 1+0+0+5 = 6$$

$$o_{1,2}: \quad \begin{pmatrix} 2 & 3 \\ 5 & 6 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix} \rightarrow o_{1,2} = 2+0+0+6 = 8$$

$$o_{2,1}: \quad \begin{pmatrix} 4 & 5 \\ 1 & 5 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 5 \end{pmatrix} \rightarrow o_{2,1} = 4+0+0+5 = 9$$

$$o_{2,2}: \quad \begin{pmatrix} 5 & 6 \\ 5 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix} \rightarrow o_{2,2} = 5+0+0+3 = 8$$

Neural Networks and Regularization

Neural Networks and Regularization

Task 3: Gradient for *LP* Regularization

Given:

$$\bar{J} = J(\boldsymbol{w}) + \alpha \|\boldsymbol{w}\|_p^p$$

$$\|\boldsymbol{w}\|_p^p = \sum_i w_i^p$$

Compute gradient $\nabla_{\boldsymbol{w}} \bar{J}$:

- Begin with

$$\nabla_{\boldsymbol{w}} \bar{J} = \nabla_{\boldsymbol{w}} J(\boldsymbol{w}) + \nabla_{\boldsymbol{w}} \left(\alpha \|\boldsymbol{w}\|_p^p \right)$$

- Then compute element-wise gradients independently

$$\begin{aligned}
\nabla_{w_i} \bar{J} &= \nabla_{w_i} J + \nabla_{w+i} \left(\alpha \|\boldsymbol{w}\|_p^p \right) \\
&= \nabla_{w_i} J + \alpha \cdot \nabla_{w+i} \sum_j w_j^p \\
&= \nabla_{w_i} J + \alpha \cdot \nabla_{w+i} w_i^p \\
&= \nabla_{w_i} J + \alpha \cdot p \cdot w_i^{p-1}
\end{aligned}$$