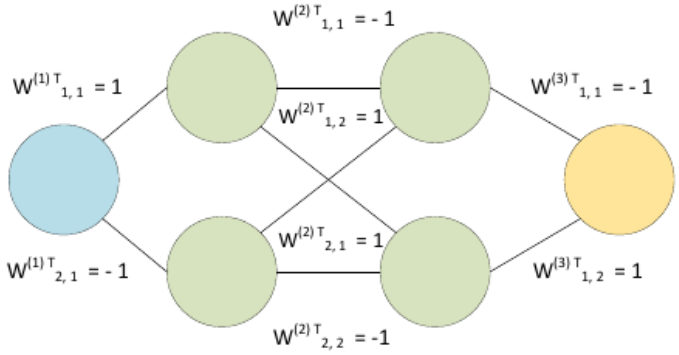
Submission Date: 6th December 2023

1 Backpropagation

- What is the advantage of backpropagation vs. computing the gradients in isolation for every parameter ?
- You are given a network with 3 layers, no bias terms and with ReLU activation functions for the hidden layers (green). The weights are initialized as shown in the figure below, the learning rate is $\eta = 0.01$ and the the mean squared error is used as a loss function. Given a dataset of only one example that has only one feature $X = \{3\}$ and label $y = \{6\}$, calculate the loss of the network and update the weights to their new values using backpropagation.



2 Convolutional Neural Network (CNN)

- What is the main structural difference between CNNs and Feed Forward Neural Networks?
- You are given a 4 x 4 grayscale image and a convolutional layer that features 2 kernels of size 2 x 2 with stride 1. What would be the output size after feeding the input image to the convolutional layer? What would be the output size if we increased the stride to a value of 2? With stride, we refer to the the sliding movement of the kernel. Do not apply any padding.
- Assume that you have an image I and a kernel K:

$$I = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 5 & 3 \end{pmatrix} \quad K = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{1}$$

Compute the convolution $O = I * K$ without using padding, i.e. the output matrix has dimensions $O \in \mathrm{R}^{2\times 2}$.

3 Neural Networks and Regularization

LP regularization penalizes the p-norm of the weights, i.e., $LP = \|\mathbf{w}\|_p^p = \sum_i w_i^p$. This regularization generalizes $L1$ and $L2$ regularization. Compute the gradients of the weights for the loss function using LP regularization:

$$\tilde{J} = J(\mathbf{w}) + \alpha\|\mathbf{w}\|_p^p \tag{2}$$

$$\nabla_{\mathbf{w}}\tilde{J} = ? \tag{3}$$