

Introduction to Neural Networks 2

CS 486/686

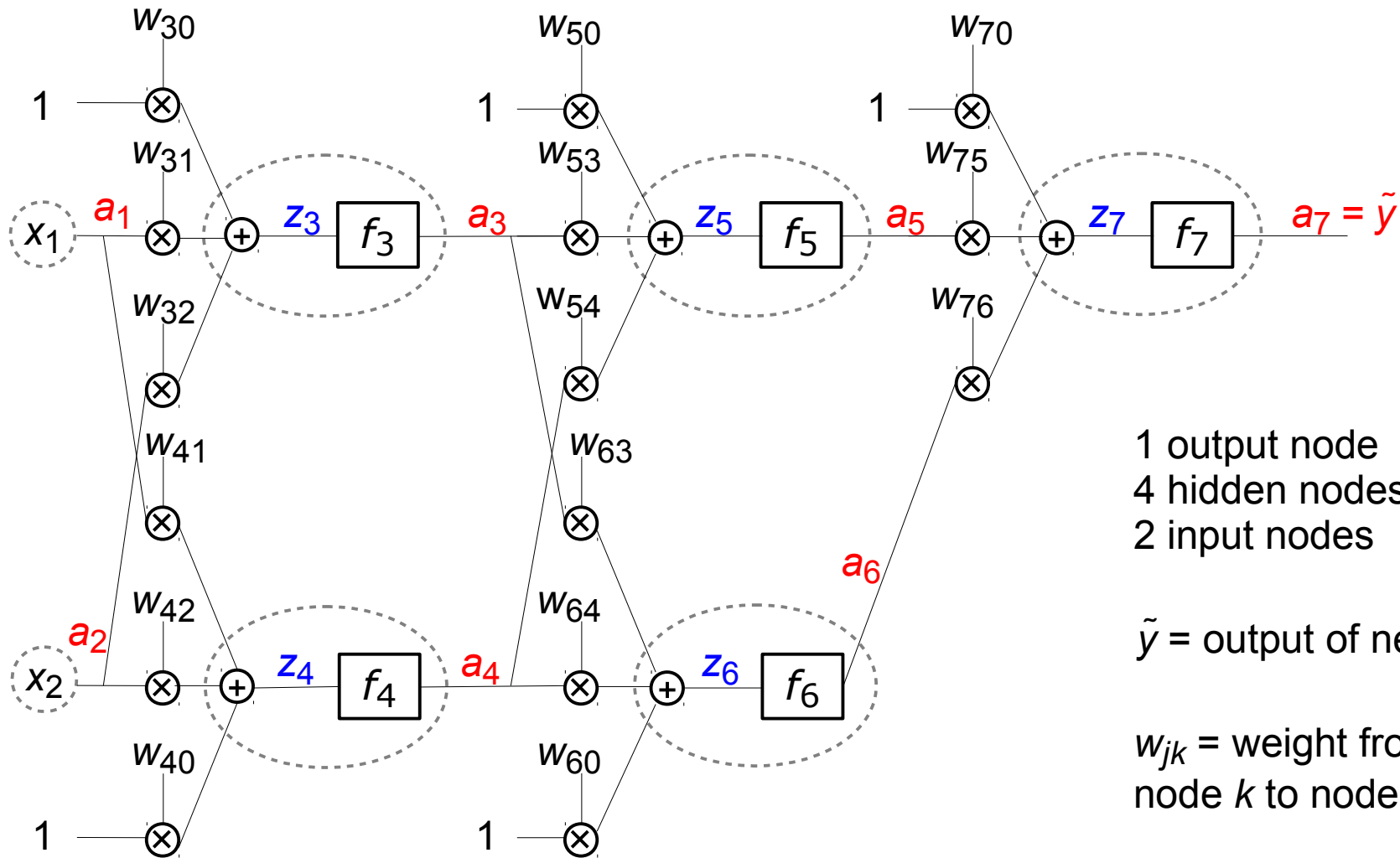
University of Waterloo

Lecture 20: July 9, 2015

Multi-Layer Neural Networks

- Perceptron can only represent linear separators
- Need multiple layers to represent more complicated separators

Example: Two Hidden Layers



Learning Multi-Layer Network

- Minimize squared error:

$$E(w, x, y) = \frac{1}{2}(y - \tilde{y})^2$$

$$\hat{w} = \arg \min_w E(w, x, y)$$

- Solution: gradient descent
- Just like what we did with sigmoid perceptron.

Learning Multi-Layer Network

- Minimize squared error:

$$E(w, x, y) = \frac{1}{2}(y - \tilde{y})^2$$

$$\hat{w} = \arg \min_w E(w, x, y)$$

- Solution: gradient descent
- Just like what we did with sigmoid perceptron.
- Problem: gradient much harder to compute

Learning Multi-Layer Network

- Minimize squared error:

$$E(w, x, y) = \frac{1}{2}(y - \tilde{y})^2$$

$$\hat{w} = \arg \min_w E(w, x, y)$$

- Solution: gradient descent
- Just like what we did with sigmoid perceptron.
- Problem: gradient much harder to compute
 - Solution: compute gradient with backpropagation

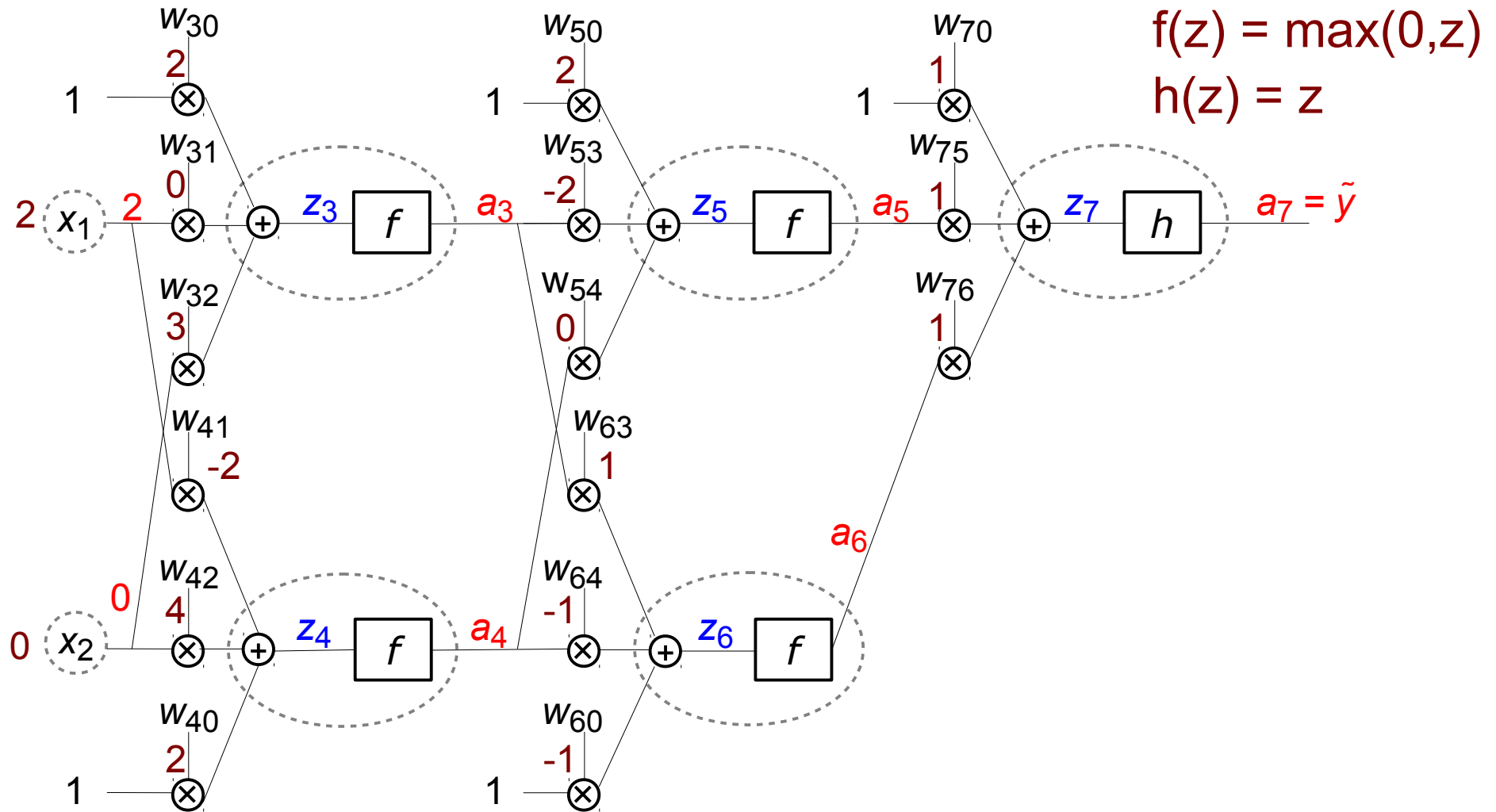
Learning Multi-Layer Network by Gradient Descent

- Initialize weights w
- For each training example (\mathbf{x}, y) do
 - Compute gradient $\nabla E(w)$ by backpropagation
 - Update weights $w \leftarrow w - \alpha \nabla E(w)$
- Repeat until stopping criteria satisfied

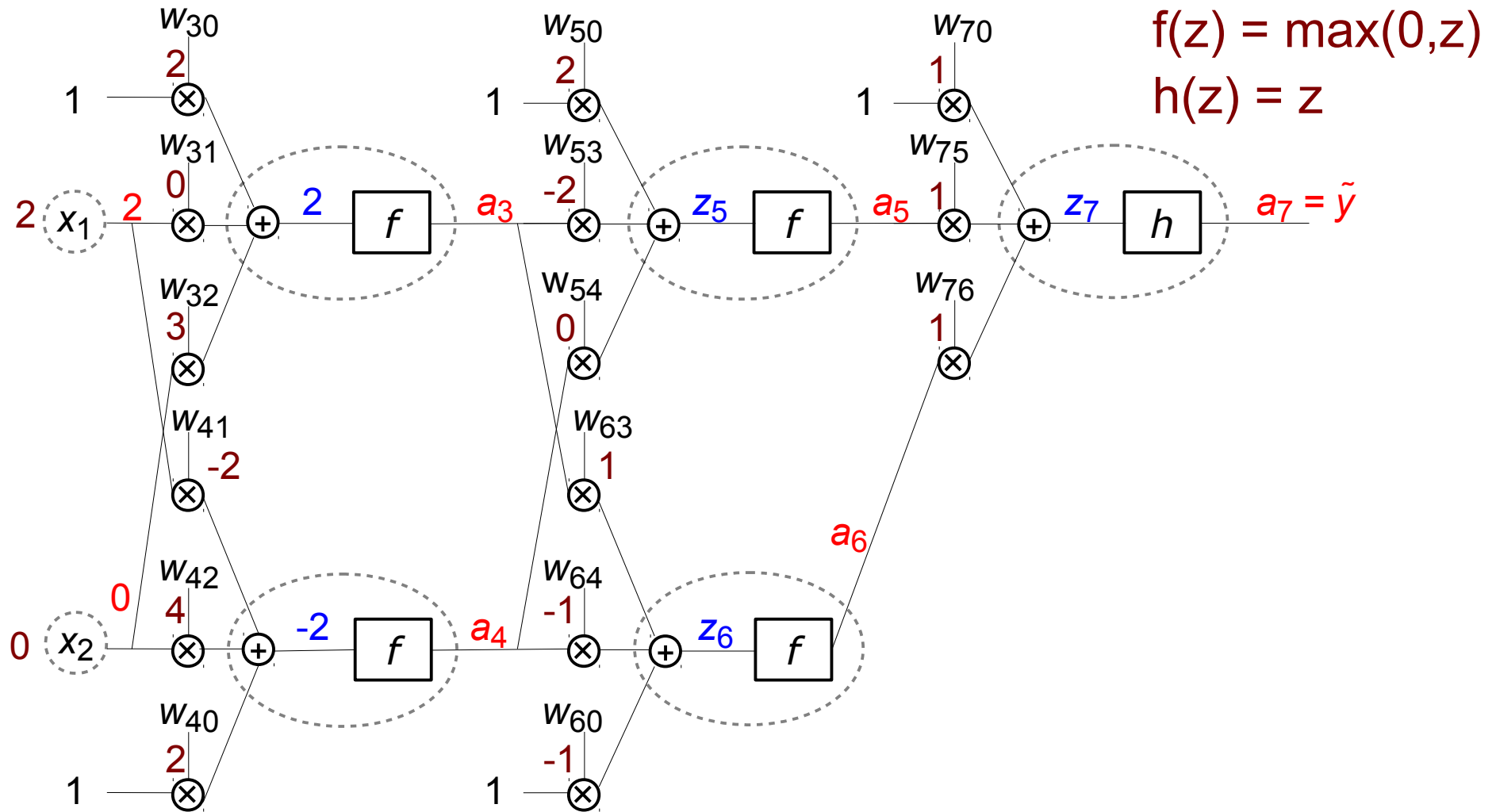
Backpropagation

- Given training example (x, y)
- Forward phase
 - Starting from input nodes, compute all z_k 's and a_k 's by forward propagation
- Backward phase
 - Starting from output nodes, compute $D_k = -\frac{\partial E}{\partial z_k}$:
 - $D_k = f'(z_k)(a_k - y_k)$ if k is an output node
 - $D_k = f'(z_k) \sum_j w_{jk} D_j$ if k is a hidden node
 - Compute $\frac{\partial E}{\partial w_{jk}} = -D_j a_k$ for all w_{jk}
- Return $\nabla E(w)$

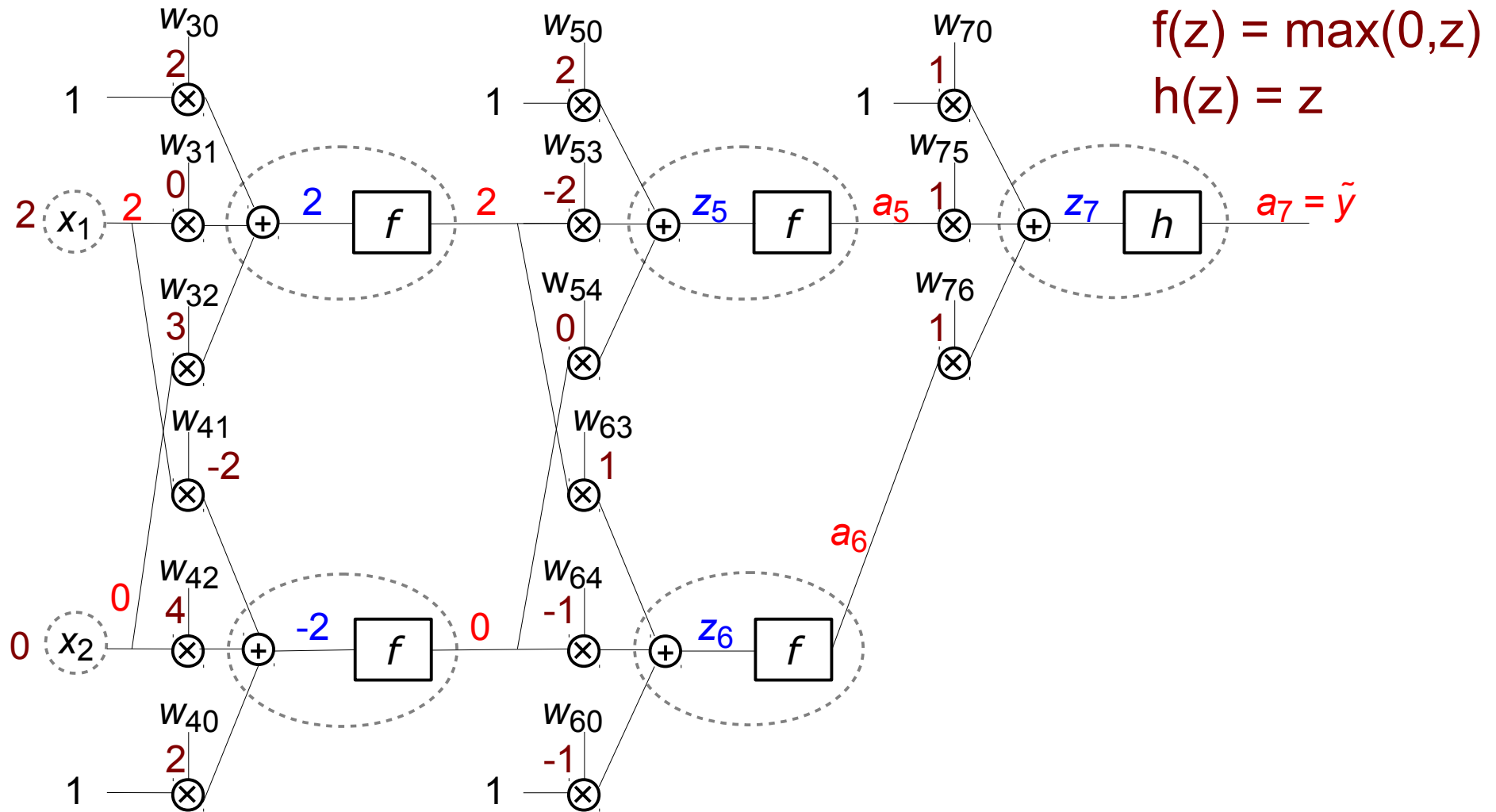
Example: Forward Phase



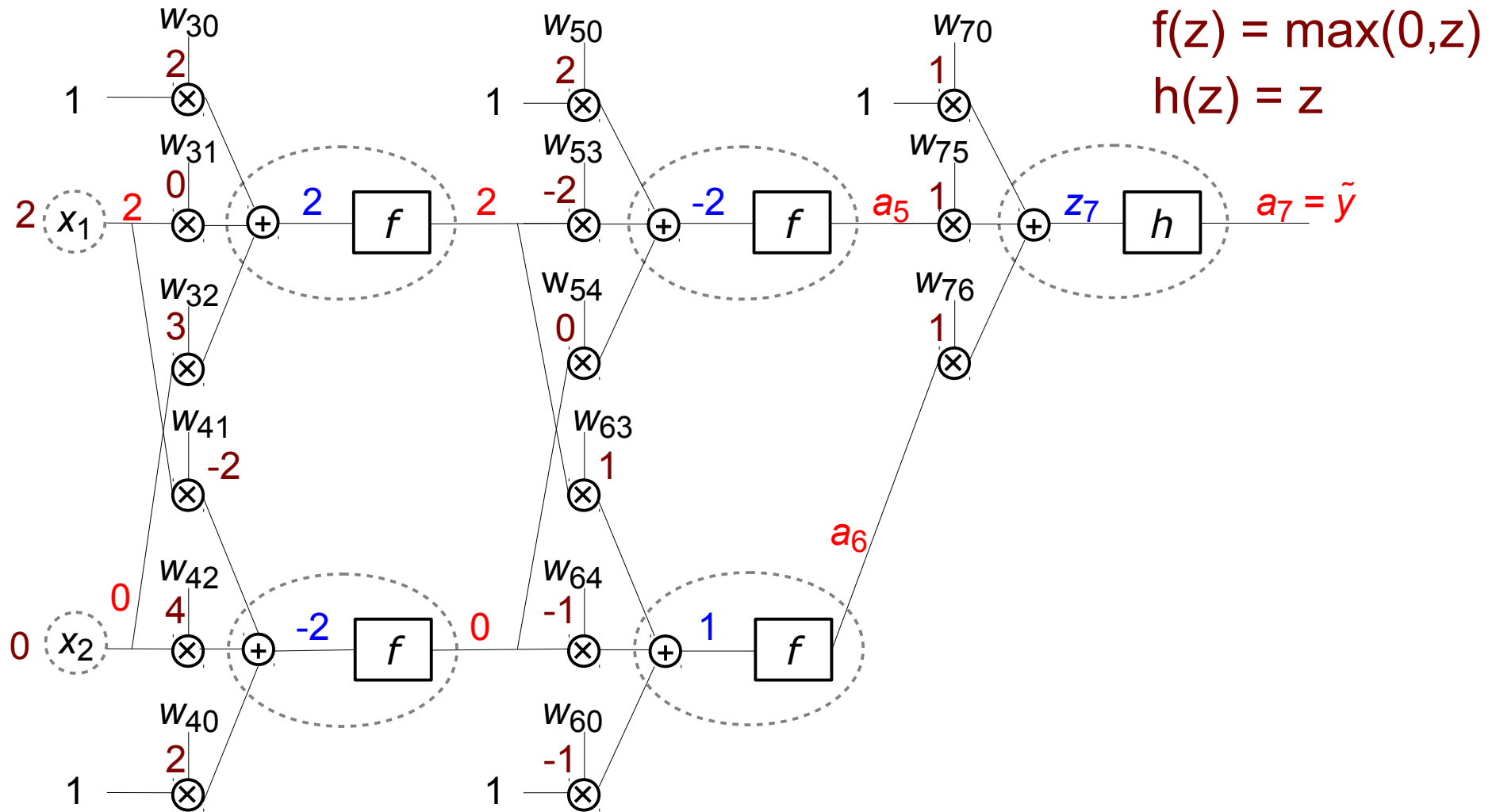
Example: Forward Phase



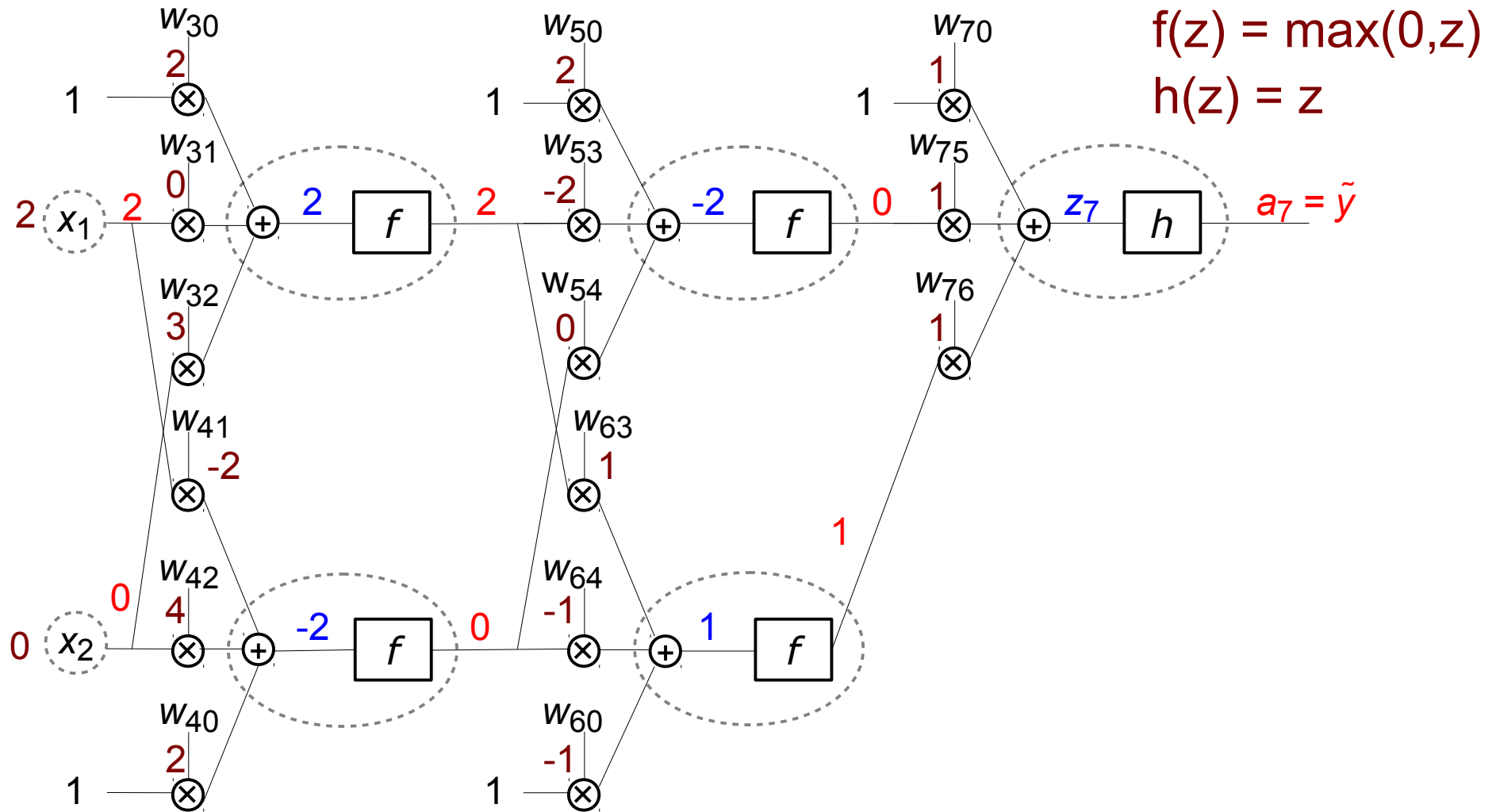
Example: Forward Phase



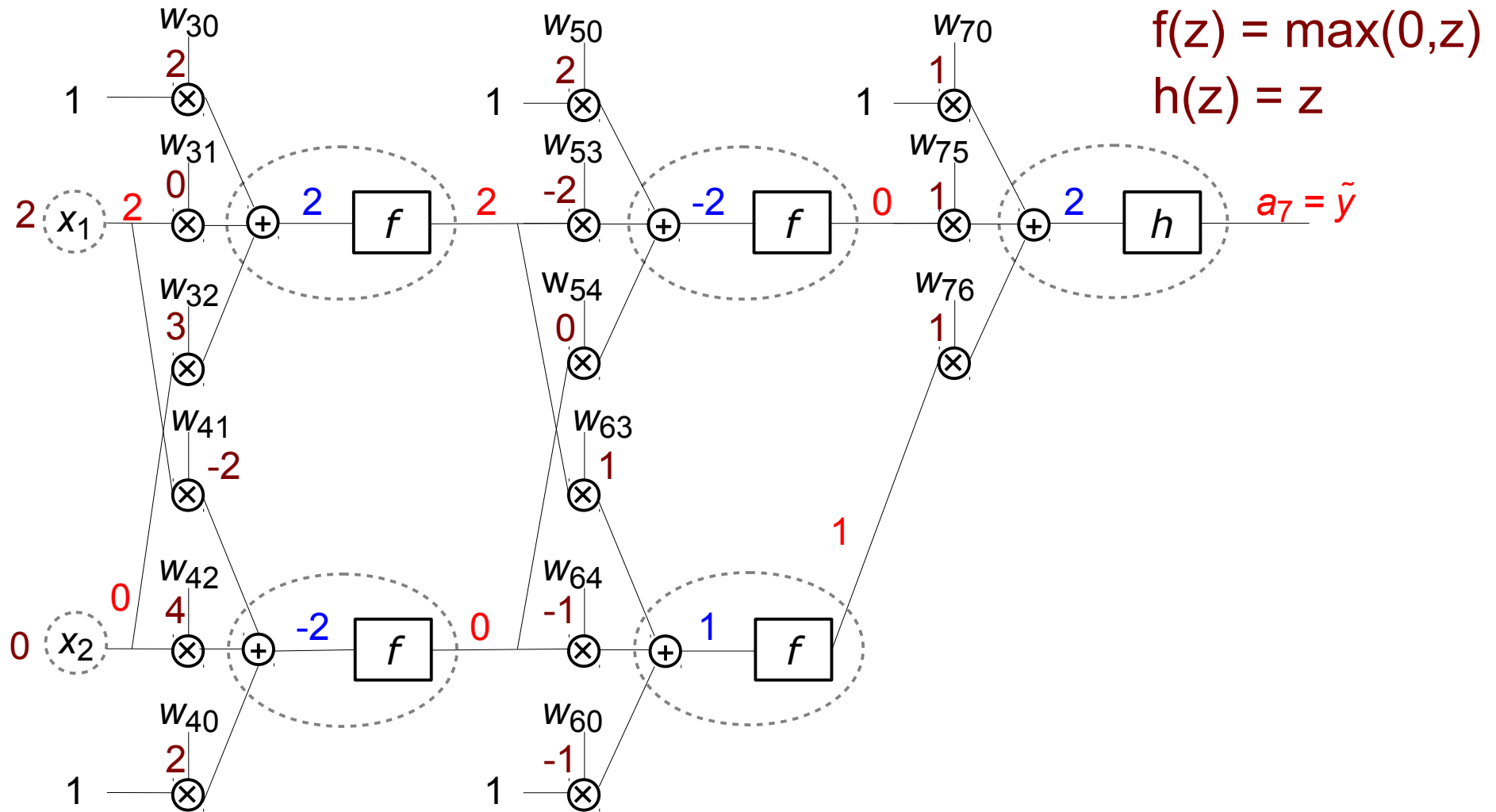
Example: Forward Phase



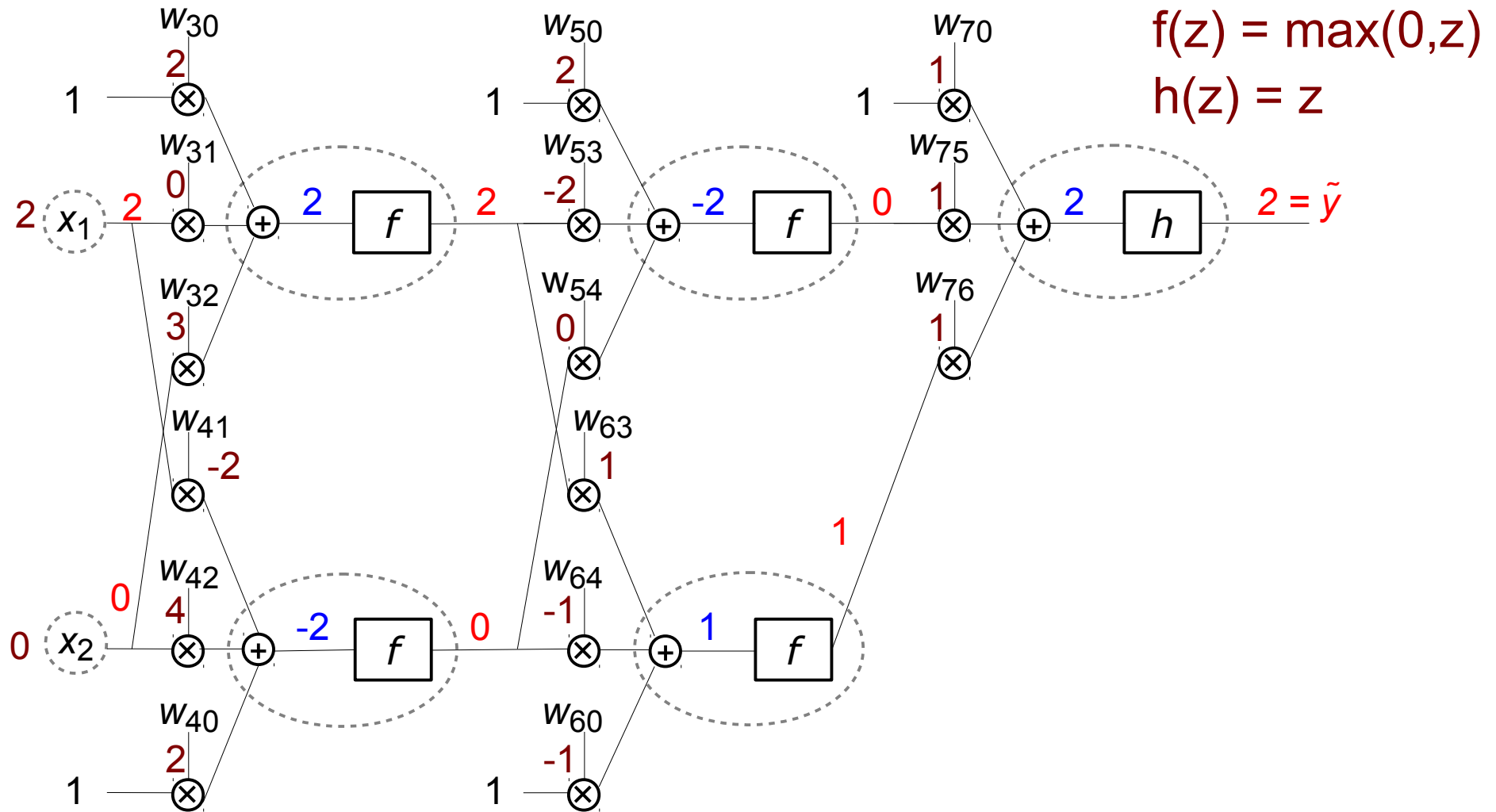
Example: Forward Phase



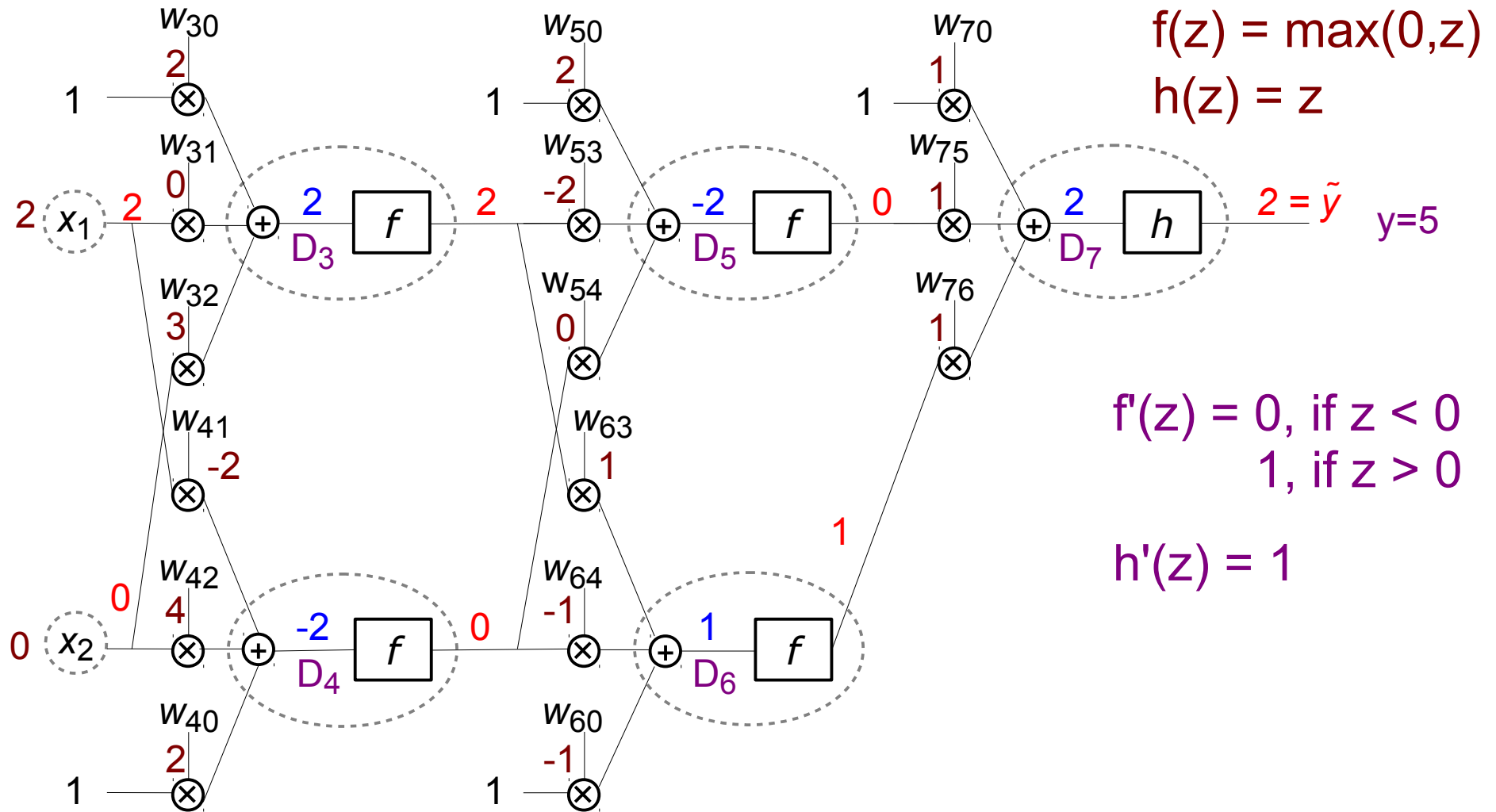
Example: Forward Phase



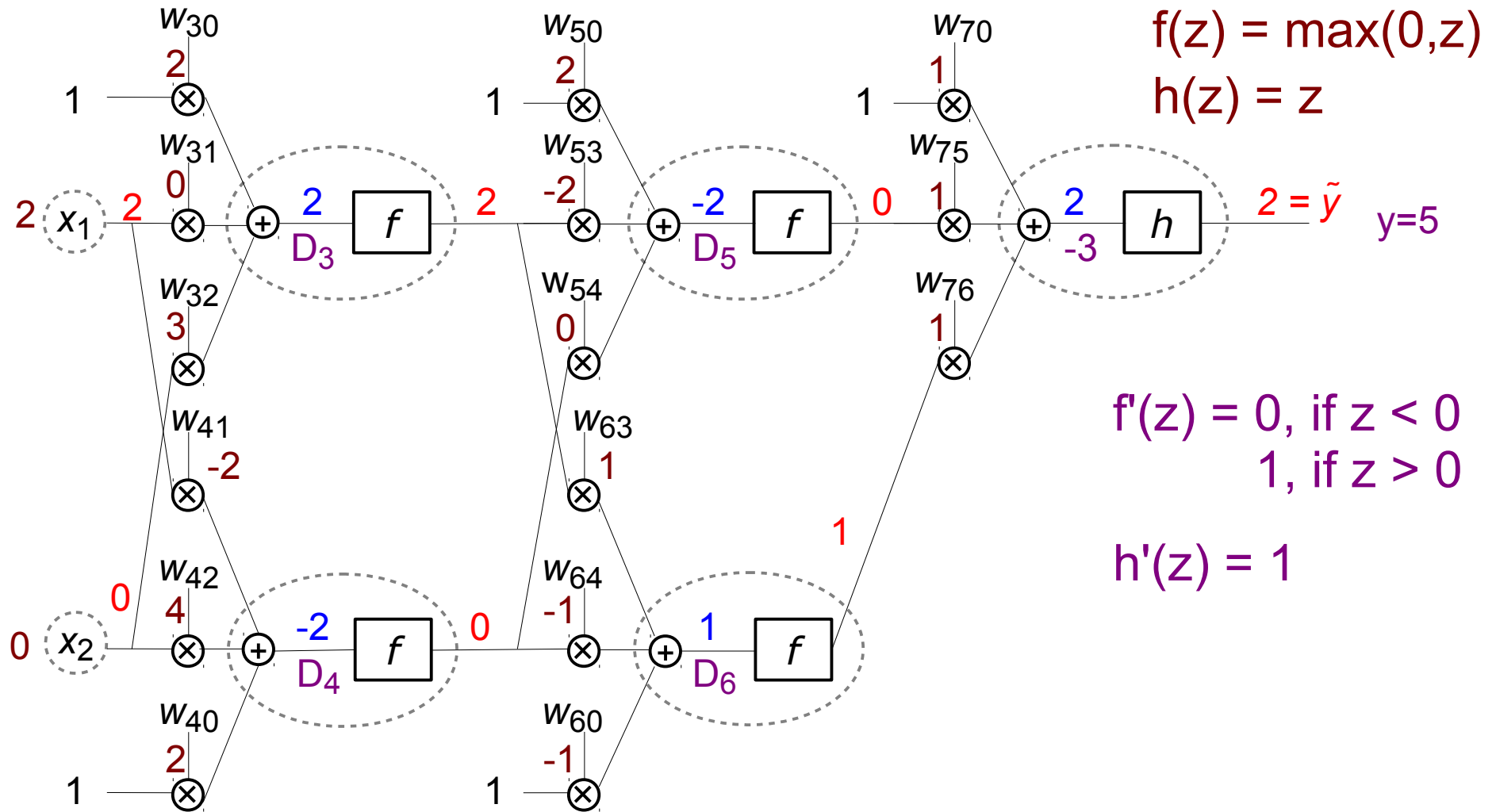
Example: Forward Phase



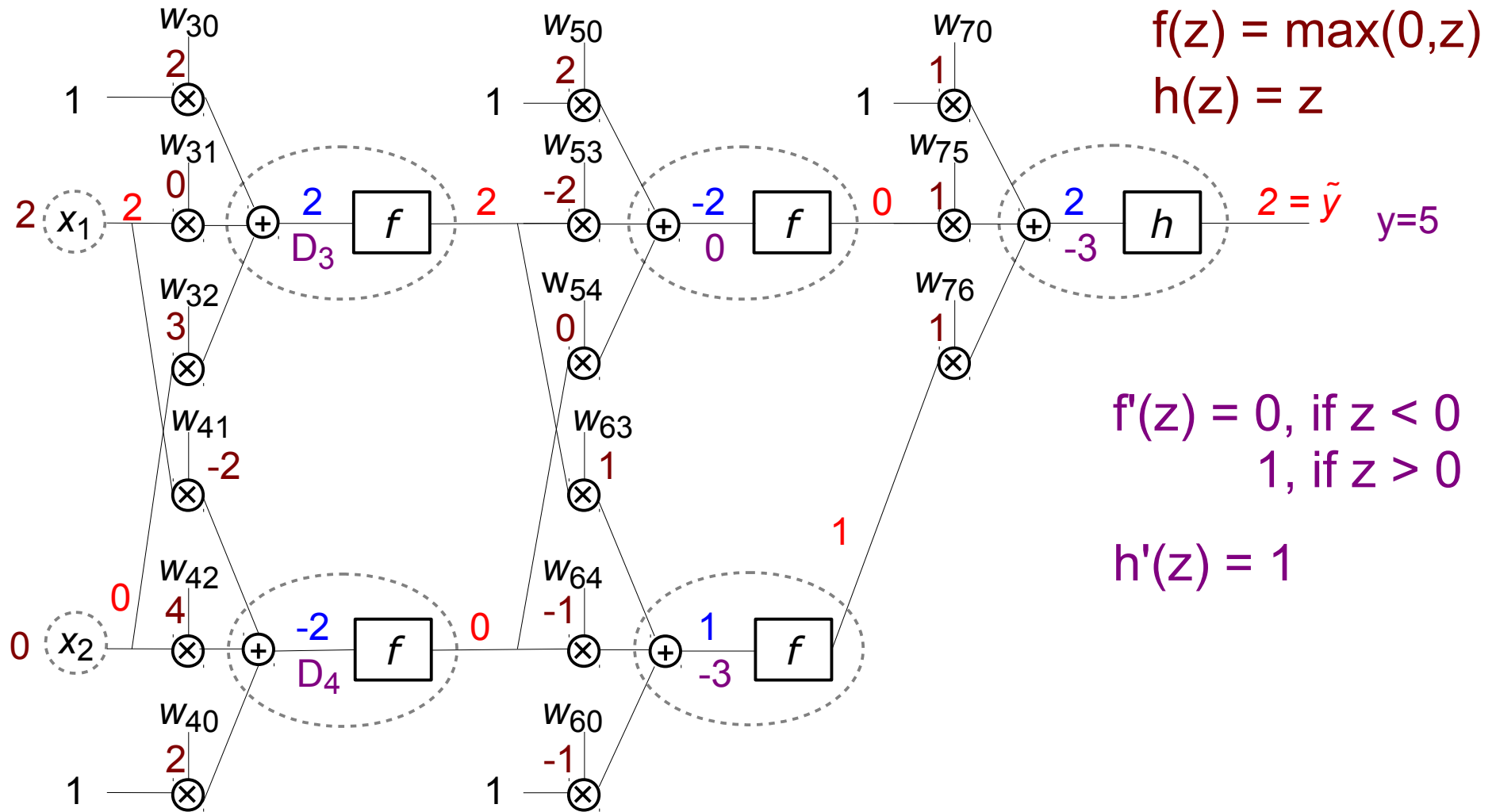
Example: Backward Phase



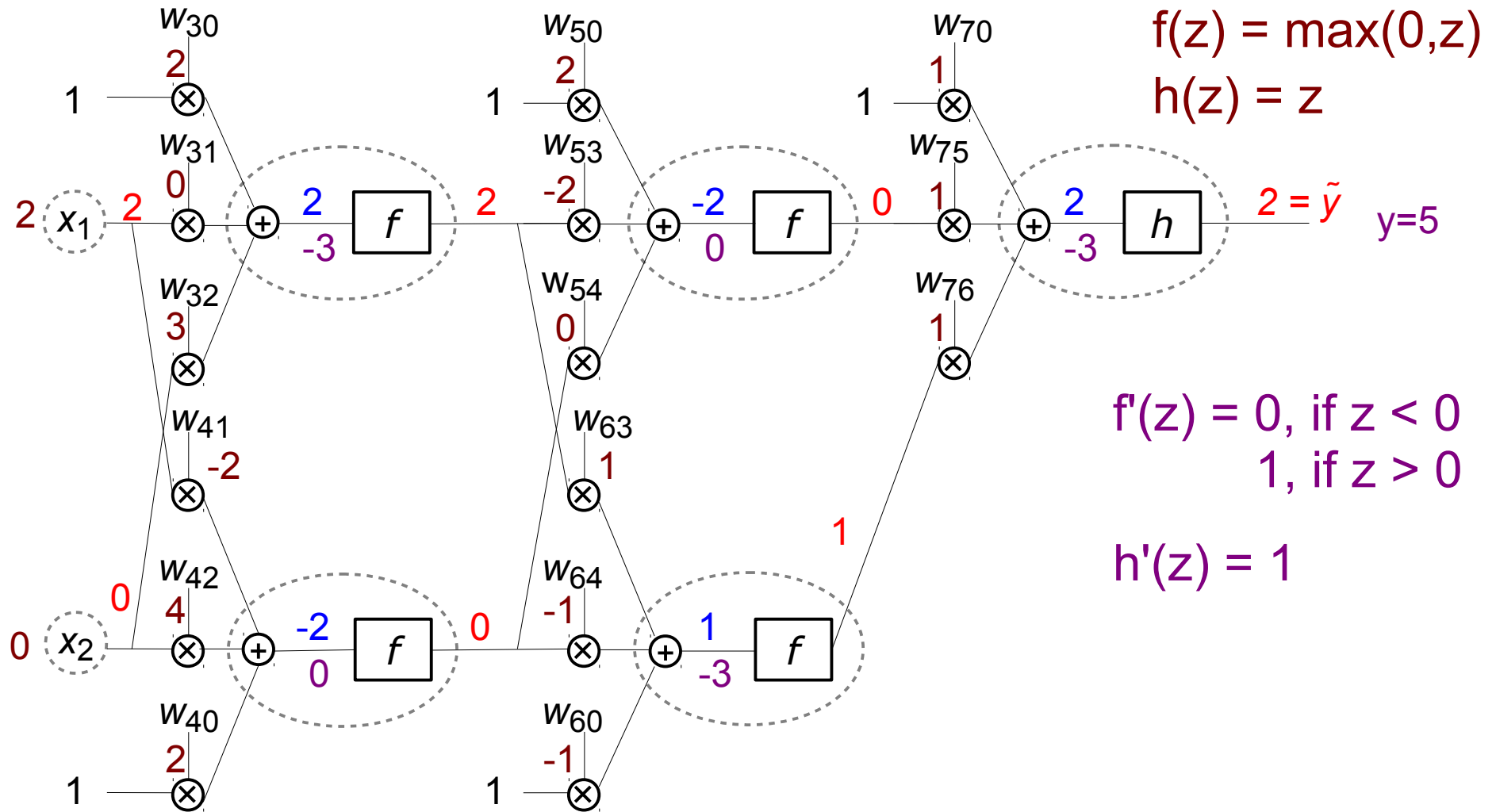
Example: Backward Phase



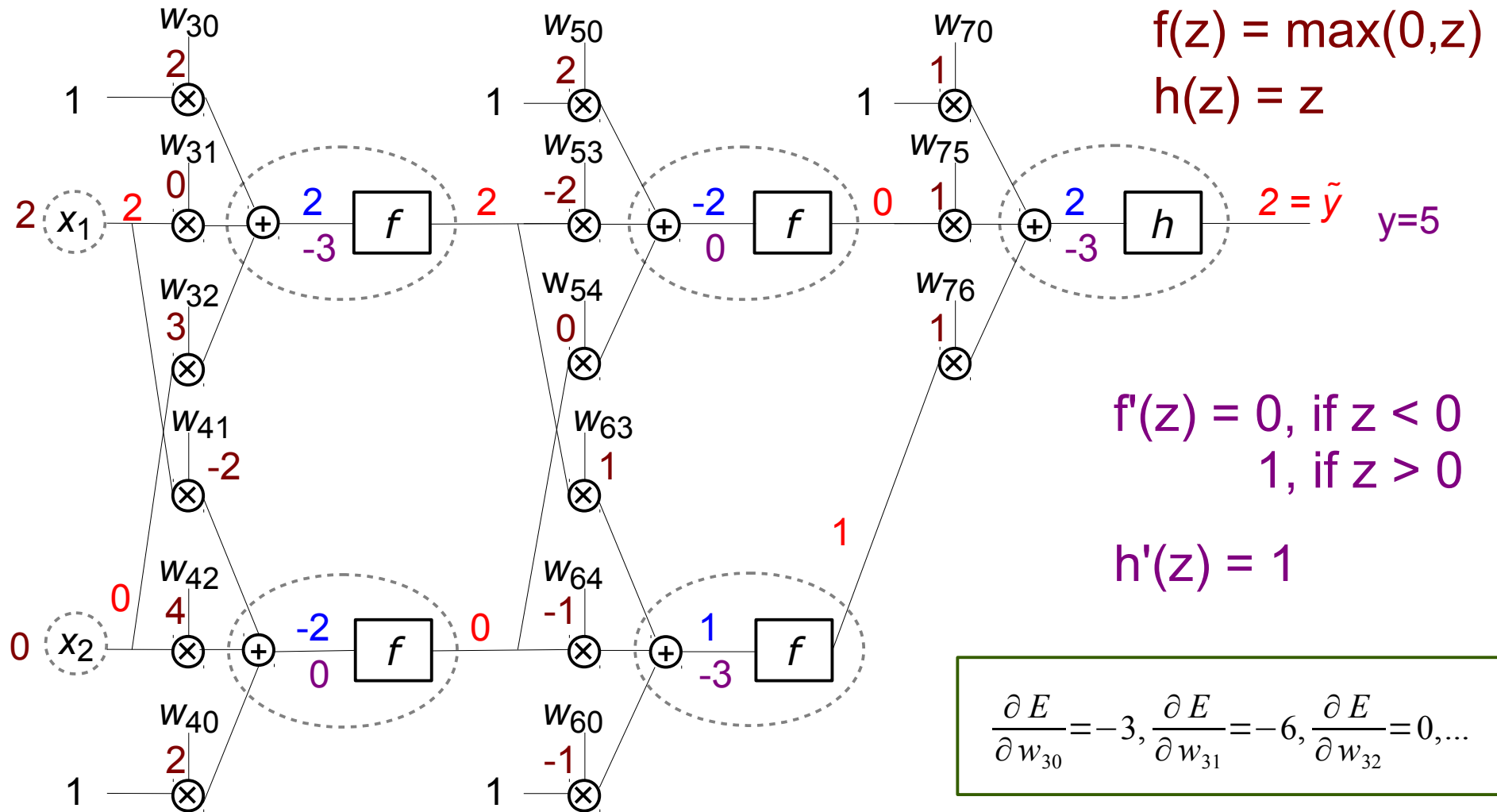
Example: Backward Phase



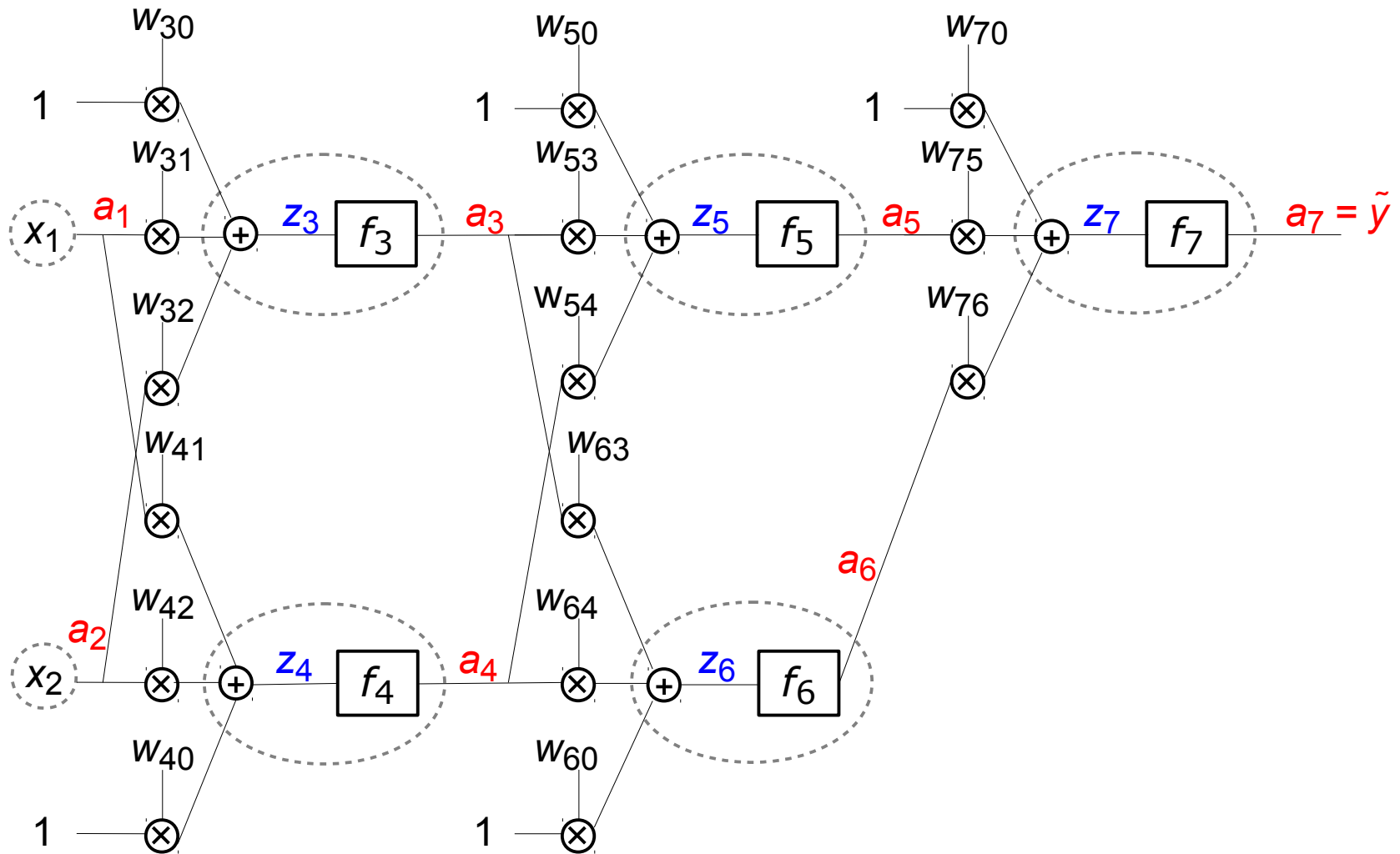
Example: Backward Phase



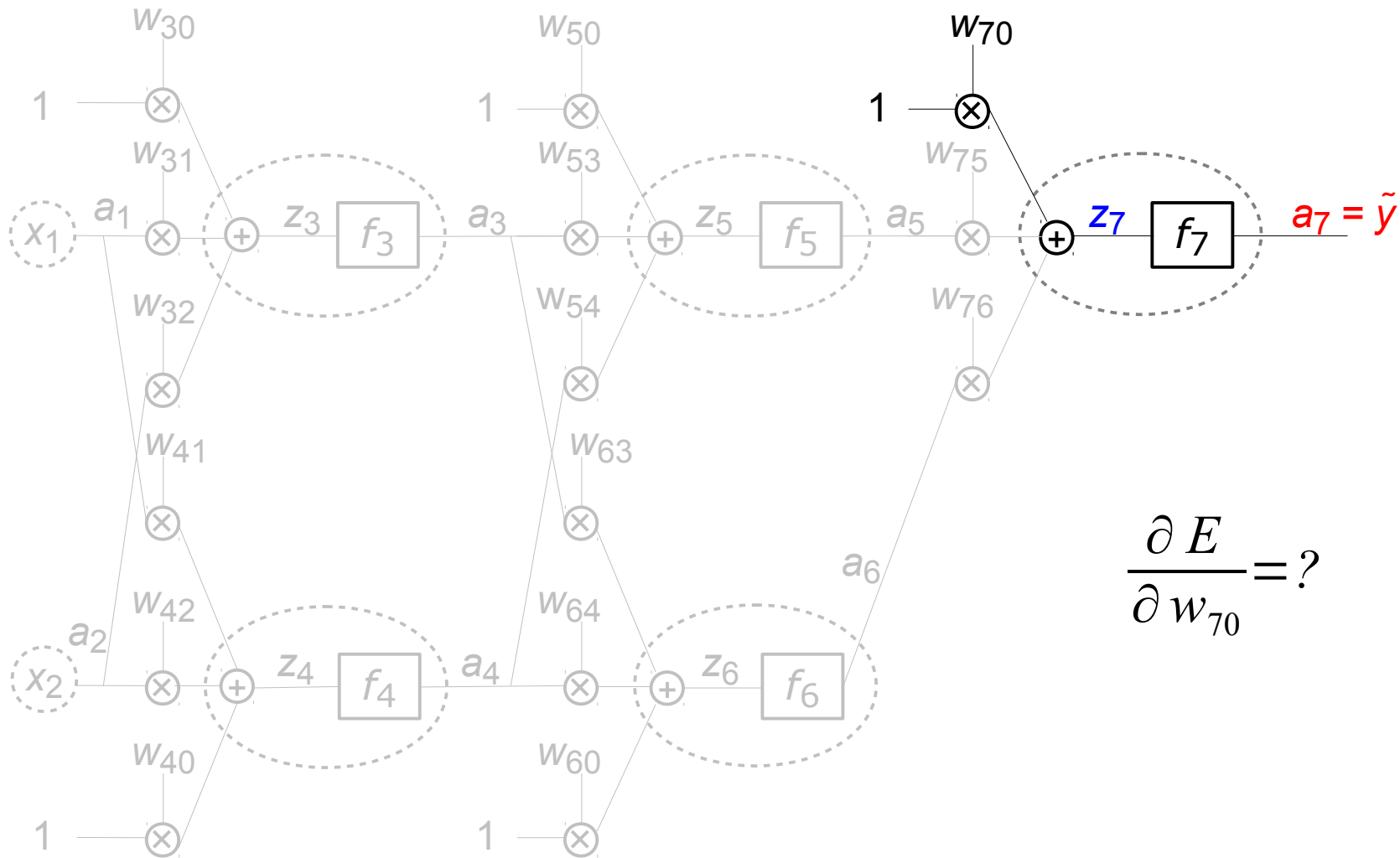
Example: Backward Phase



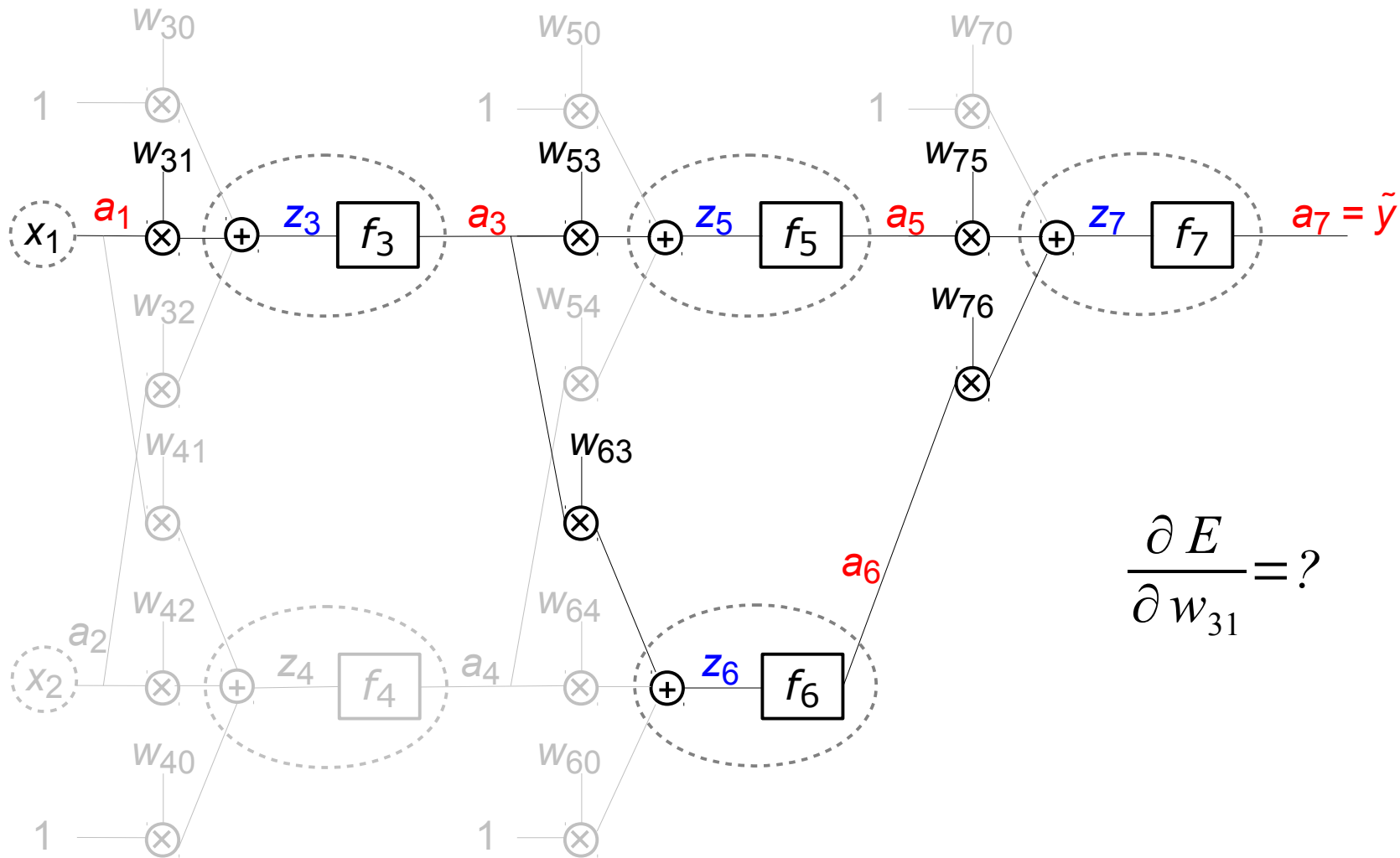
Backpropagation Derivation



Backpropagation Derivation



Backpropagation Derivation



$$\frac{\partial E}{\partial w_{31}} = ?$$