Backpropagation: A complete example

Lorenzo Mannocci

a.y. 2026/2027

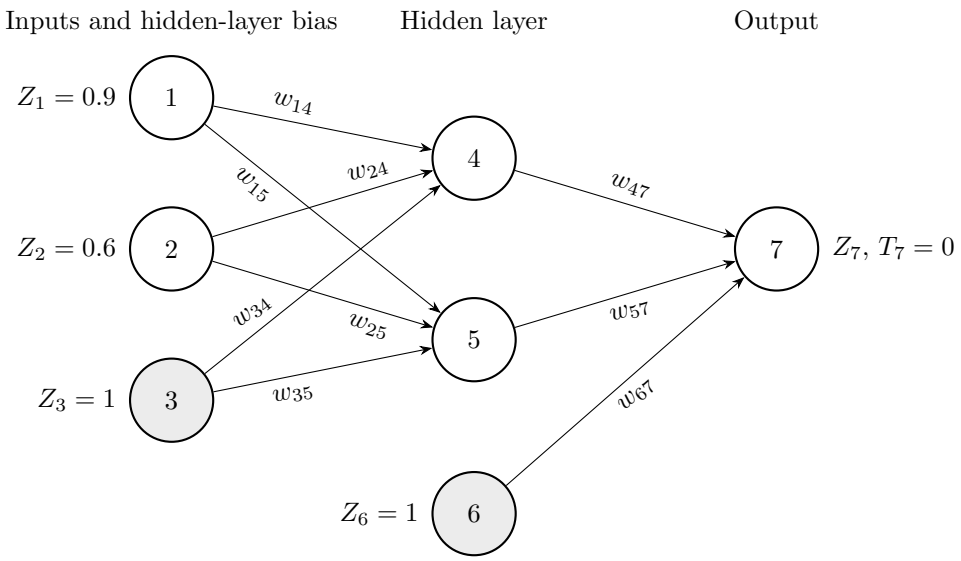
1 Exercise and input parameters

Consider a feedforward neural network with two input features, two sigmoid hidden neurons, and one sigmoid output neuron. Perform one forward pass and one backpropagation update for the given training example.

- Input values: $Z_1 = x_1 = 0.9$ and $Z_2 = x_2 = 0.6$.
- Constant bias inputs: $Z_3 = 1$ for hidden neurons 4 and 5, and $Z_6 = 1$ for output neuron 7.
- Hidden neurons: 4 and 5. Output neuron: 7.
- Target: $T_7 = 0$.
- Learning rate: $\lambda = 1$.
- All nine connection weights are initially 0.5:

$$w_{14} = w_{24} = w_{34} = w_{15} = w_{25} = w_{35} = w_{47} = w_{57} = w_{67} = 0.5.$$

The constant nodes 3 and 6 do not apply a sigmoid and do not have error signals to compute. Their outgoing weights are trainable bias parameters. In this example bias contributions are *added* to the weighted sums.



2 Theoretical formulas

For a connection $i \to j$, w_{ij} denotes the weight from source neuron i to destination neuron j. Let $\mathcal{P}_j$ be the set of inputs to neuron j, and let Φ_j be its downstream neurons.

Forward pass and activation derivative

$$net_j = \sum_{i \in \mathcal{P}_j} w_{ij} Z_i, \qquad Z_j = f(net_j) = \frac{1}{1 + e^{-net_j}}, \qquad f'(net_j) = Z_j(1 - Z_j).$$

The formulas below correspond to the squared-error loss for one training example:

$$E = \frac{1}{2} \sum_{j \in \text{output}} (T_j - Z_j)^2.$$

Output-neuron error signal

$$\boxed{\delta_j = (T_j - Z_j) f'(net_j) = (T_j - Z_j) Z_j (1 - Z_j).}$$

Hidden-neuron error signal

$$\boxed{\delta_j = f'(net_j) \sum_{k \in \Phi_j} \delta_k w_{jk} = Z_j(1 - Z_j) \sum_{k \in \Phi_j} \delta_k w_{jk}.}$$

Weight variation and new weight

$$\boxed{\Delta w_{ij} = \lambda \delta_j Z_i, \qquad w_{ij}^{\text{new}} = w_{ij}^{\text{old}} + \Delta w_{ij}.}$$

The factor is Z_i, the *source* output, not Z_j. For a bias connection, $Z_i = 1$, so $\Delta w_{ij} = \lambda \delta_j$. With this definition of δ_j, the plus sign in the update implements gradient descent.

Order of computation: first compute all activations, then the output error signal, then both hidden error signals, and finally update all weights. All error signals use the original weights from the forward pass.

3 Step 1: Forward pass

Input and bias nodes 1, 2, 3, and 6

These outputs are supplied directly:

$$Z_1 = 0.9, \qquad Z_2 = 0.6, \qquad Z_3 = 1, \qquad Z_6 = 1.$$

Hidden neuron 4

$$\begin{aligned} net_4 &= w_{14} Z_1 + w_{24} Z_2 + w_{34} Z_3 \\ &= (0.5)(0.9) + (0.5)(0.6) + (0.5)(1) = 1.25, \\ Z_4 &= \frac{1}{1 + e^{-net_4}} = \frac{1}{1 + e^{-1.25}} \approx 0.77, \end{aligned}$$

$$f'(net_4) = Z_4(1 - Z_4) \approx (0.77)(1 - 0.77) \approx 0.17.$$

Hidden neuron 5

$$\begin{aligned} net_5 &= w_{15}Z_1 + w_{25}Z_2 + w_{35}Z_3 \\ &= (0.5)(0.9) + (0.5)(0.6) + (0.5)(1) = 1.25, \\ Z_5 &= \frac{1}{1 + e^{-net_5}} = \frac{1}{1 + e^{-1.25}} \approx 0.77, \\ f'(net_5) &= Z_5(1 - Z_5) \approx (0.77)(1 - 0.77) \approx 0.17. \end{aligned}$$

Output neuron 7

$$\begin{aligned} net_7 &= w_{47}Z_4 + w_{57}Z_5 + w_{67}Z_6 \\ &\approx (0.5)(0.77) + (0.5)(0.77) + (0.5)(1) \\ &\approx 1.27, \\ Z_7 &= \frac{1}{1 + e^{-net_7}} \approx \frac{1}{1 + e^{-1.27}} \approx 0.78, \\ f'(net_7) &= Z_7(1 - Z_7) \approx (0.78)(1 - 0.78) \approx 0.17. \end{aligned}$$

The output exceeds the target $T_7 = 0$, so training should reduce it.

4 Step 2: Backpropagate the error signals

Output neuron 7

$$\begin{aligned} \delta_7 &= (T_7 - Z_7)f'(net_7) = (T_7 - Z_7)Z_7(1 - Z_7) \\ &\approx (0 - 0.78)(0.78)(1 - 0.78) \\ &\approx \boxed{-0.13}. \end{aligned}$$

Hidden neuron 4

Neuron 4 feeds only output neuron 7: $\Phi_4 = \{7\}$.

$$\begin{aligned} \delta_4 &= Z_4(1 - Z_4) \sum_{k \in \Phi_4} \delta_k w_{4k} \\ &= Z_4(1 - Z_4)\delta_7 w_{47} \\ &\approx (0.77)(1 - 0.77)(-0.13)(0.5) \\ &\approx \boxed{-0.01}. \end{aligned}$$

Hidden neuron 5

Neuron 5 feeds only output neuron 7: $\Phi_5 = \{7\}$.

$$\delta_5 = Z_5(1 - Z_5) \sum_{k \in \Phi_5} \delta_k w_{5k}$$

$$
\begin{aligned}
&= Z_5(1 - Z_5)\delta_7 w_{57} \\
&\approx (0.77)(1 - 0.77)(-0.13)(0.5) \\
&\approx \boxed{-0.01}.
\end{aligned}
$$

The two hidden signals are equal because the neurons have identical initial incoming and outgoing weights.

5 Step 3: Update every connection weight

Each calculation uses $\Delta w_{ij} = \lambda \delta_j Z_i$ and $w_{ij}^{\text{new}} = w_{ij}^{\text{old}} + \Delta w_{ij}$.

Weights entering hidden neuron 4

Connection $1 \to 4$.

$$
\begin{aligned}
\Delta w_{14} &= \lambda \delta_4 Z_1 \approx (1)(-0.01)(0.9) = -0.01, \\
w_{14}^{\text{new}} &= w_{14}^{\text{old}} + \Delta w_{14} \approx 0.5 - 0.01 = \boxed{0.49}.
\end{aligned}
$$

Connection $2 \to 4$.

$$
\begin{aligned}
\Delta w_{24} &= \lambda \delta_4 Z_2 \approx (1)(-0.01)(0.6) = -0.006, \\
w_{24}^{\text{new}} &= w_{24}^{\text{old}} + \Delta w_{24} \approx 0.5 - 0.006 = \boxed{0.494}.
\end{aligned}
$$

Bias connection $3 \to 4$.

$$
\begin{aligned}
\Delta w_{34} &= \lambda \delta_4 Z_3 \approx (1)(-0.01)(1) = -0.01, \\
w_{34}^{\text{new}} &= w_{34}^{\text{old}} + \Delta w_{34} \approx 0.5 - 0.01 = \boxed{0.49}.
\end{aligned}
$$

Weights entering hidden neuron 5

Connection $1 \to 5$.

$$
\begin{aligned}
\Delta w_{15} &= \lambda \delta_5 Z_1 \approx (1)(-0.01)(0.9) = -0.01, \\
w_{15}^{\text{new}} &= w_{15}^{\text{old}} + \Delta w_{15} \approx 0.5 - 0.01 = \boxed{0.49}.
\end{aligned}
$$

Connection $2 \to 5$.

$$
\begin{aligned}
\Delta w_{25} &= \lambda \delta_5 Z_2 \approx (1)(-0.01)(0.6) = -0.006, \\
w_{25}^{\text{new}} &= w_{25}^{\text{old}} + \Delta w_{25} \approx 0.5 - 0.006 = \boxed{0.494}.
\end{aligned}
$$

Bias connection $3 \to 5$.

$$
\begin{aligned}
\Delta w_{35} &= \lambda \delta_5 Z_3 \approx (1)(-0.01)(1) = -0.01, \\
w_{35}^{\text{new}} &= w_{35}^{\text{old}} + \Delta w_{35} \approx 0.5 - 0.01 = \boxed{0.49}.
\end{aligned}
$$

Weights entering output neuron 7

Connection $4 \to 7$.

$$\Delta w_{47} = \lambda \delta_7 Z_4 \approx (1)(-0.13)(0.77) = -0.1,$$
$$w_{47}^{\text{new}} = w_{47}^{\text{old}} + \Delta w_{47} \approx 0.5 - 0.1 = \boxed{0.4}.$$

Connection $5 \to 7$.

$$\Delta w_{57} = \lambda \delta_7 Z_5 \approx (1)(-0.13)(0.77) = -0.1$$
$$w_{57}^{\text{new}} = w_{57}^{\text{old}} + \Delta w_{57} \approx 0.5 - 0.1 = \boxed{0.4}.$$

Bias connection $6 \to 7$.

$$\Delta w_{67} = \lambda \delta_7 Z_6 \approx (1)(-0.13)(1) = -0.13,$$
$$w_{67}^{\text{new}} = w_{67}^{\text{old}} + \Delta w_{67} \approx 0.5 - 0.13 = \boxed{0.37}.$$

This completes one training update for the supplied example. The activations above belong to the initial forward pass; a subsequent pass would use the new weights.