Let's say we have a neural network with two layers and one output unit. The input is x, the hidden layer activation is h, and the output is y. The weights are W_1 and W_2, and the biases are b_1 and b_2. The activation function is the sigmoid function $\sigma(z) = \frac{1}{1+e^{-z}}$. The loss function is the mean squared error $L(y,t) = \frac{1}{2}(y-t)^2$, where t is the target value.

The forward pass of the network is:

$$
\begin{aligned}
h &= \sigma(W_1 x + b_1) \\
y &= \sigma(W_2 h + b_2)
\end{aligned}
$$

The backward pass of the network is:

$$
\begin{aligned}
\frac{\partial L}{\partial y} &= y - t \\
\frac{\partial L}{\partial W_2} &= \frac{\partial L}{\partial y}\frac{\partial y}{\partial W_2} \\
&= \frac{\partial L}{\partial y}\frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial W_2} \\
&= (y-t)y(1-y)h \\
\frac{\partial L}{\partial b_2} &= \frac{\partial L}{\partial y}\frac{\partial y}{\partial b_2} \\
&= \frac{\partial L}{\partial y}\frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial b_2} \\
&= (y-t)y(1-y) \\
\frac{\partial L}{\partial h} &= \frac{\partial L}{\partial y}\frac{\partial y}{\partial h} \\
&= \frac{\partial L}{\partial y}\frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial h} \\
&= (y-t)y(1-y)W_2 \\
\frac{\partial L}{\partial W_1} &= \frac{\partial L}{\partial h}\frac{\partial h}{\partial W_1} \\
&= \frac{\partial L}{\partial h}\frac{\partial h}{\partial z_1}\frac{\partial z_1}{\partial W_1} \\
&= (y-t)y(1-y)W_2 h(1-h)x \\
\frac{\partial L}{\partial b_1} &= \frac{\partial L}{\partial h}\frac{\partial h}{\partial b_1} \\
&= \frac{\partial L}{\partial h}\frac{\partial h}{\partial z_1}\frac{\partial z_1}{\partial b_1} \\
&= (y-t)y(1-y)W_2 h(1-h)
\end{aligned}
$$

where $z_1 = W_1 x + b_1$ and $z_2 = W_2 h + b_2$ are the pre-activation values of the hidden and output layers, respectively.