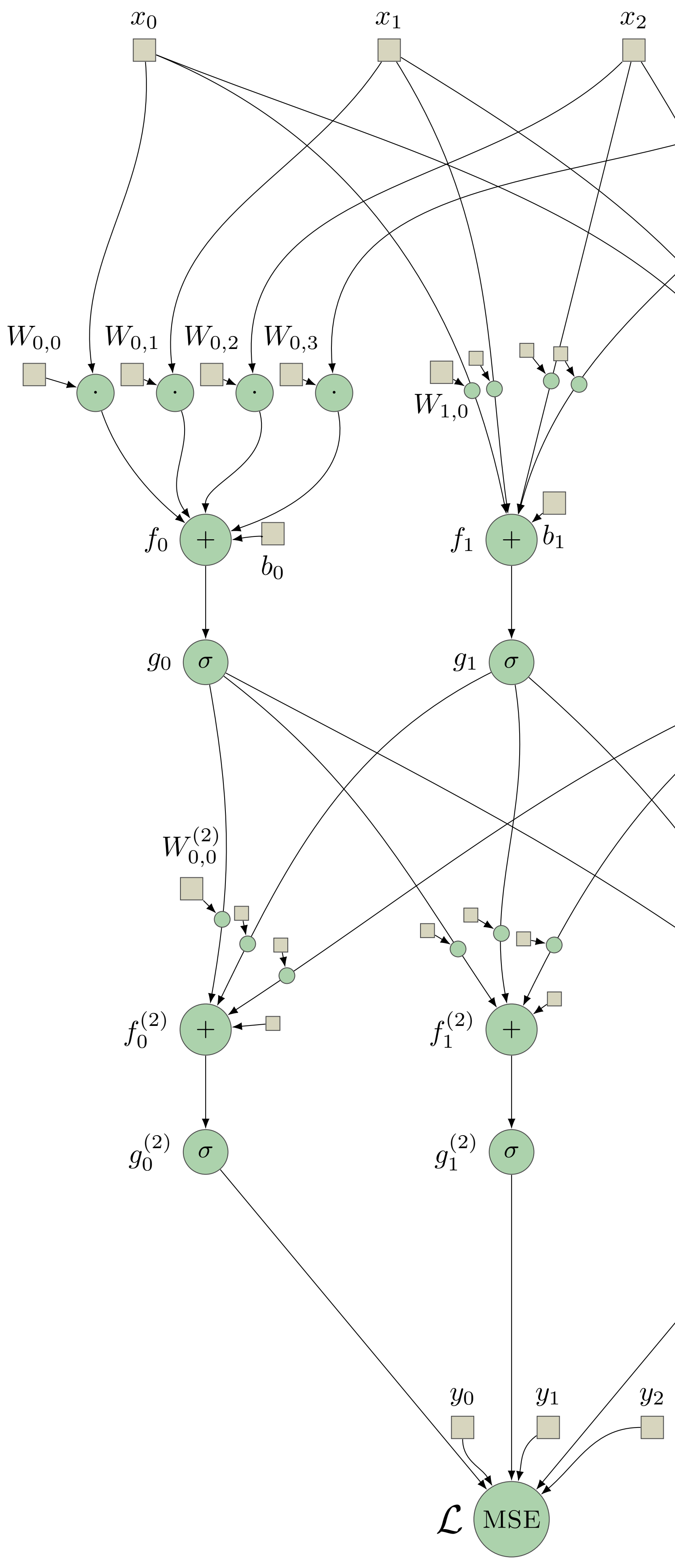

	forward (as vectors)	backward pass	backward (as vectors)
inputs $x_0, \dots, x_{N^{(0)}-1}$	$\bar{x} \in \mathbb{R}^{N^{(0)}}$		
	$\bar{W}^{(1)} \in \mathbb{R}^{N^{(1)} \times N^{(0)}}$ $\bar{b}^{(1)} \in \mathbb{R}^{N^{(1)}}$	$\frac{\partial \mathcal{L}}{\partial b_i} = \frac{\partial \mathcal{L}}{\partial f_i}$ $\frac{\partial \mathcal{L}}{\partial W_{i,j}} = \frac{\partial \mathcal{L}}{\partial f_i} \cdot x_j$	$\nabla_{\bar{b}} \mathcal{L} = \nabla_{\bar{f}^{(1)}} \mathcal{L} \quad \in \mathbb{R}^{N^{(1)}}$ $\nabla_W \mathcal{L} = (\nabla_{\bar{f}^{(1)}} \mathcal{L})\ \bar{x}^\top \quad \in \mathbb{R}^{N^{(1)} \times N^{(0)}}$
pre-activations $f_i = \sum_j W_{i,j} x_j \ + \ b_i$	$\bar{f}^{(1)} \in \mathbb{R}^{N^{(1)}}$ $\bar{f}^{(1)} = W^{(1)} \bar{x} + \bar{b}^{(1)}$		$\nabla_{\bar{f}^{(1)}} \mathcal{L} = (\nabla_{\bar{g}^{(1)}} \mathcal{L}) \odot \bar{g}^{(1)} \odot (1 - \bar{g}^{(1)})$
activations $g_i = \sigma(f_i)$	$\bar{g}^{(1)} \in \mathbb{R}^{N^{(1)}}$ $\bar{g}^{(1)} = \sigma(\bar{f}^{(1)})$	$\frac{\partial \mathcal{L}}{\partial g^{(1)}_j} = \sum_i \frac{\partial \mathcal{L}}{\partial f^{(2)}_i} \cdot W^{(2)}_{i,j}$	$\nabla_{\bar{g}^{(1)}} \mathcal{L} = W^{(2)\top} (\nabla_{\bar{f}^{(2)}} \mathcal{L})$
pre-activations$^{(2)}$ $f^{(2)}_i = \sum_j W^{(2)}_{i,j} g_j \ + \ b^{(2)}_i$	$\bar{f}^{(2)} = W^{(2)} \bar{g}^{(1)} + \bar{b}^{(2)}$	$\frac{\partial \mathcal{L}}{\partial f^{(2)}_i} = \frac{\partial \mathcal{L}}{\partial g^{(2)}_i} \cdot \sigma'(f^{(2)}_i)$ where $\sigma' = \sigma \cdot (1 - \sigma)$	$\nabla_{\bar{f}^{(2)}} \mathcal{L} = (\nabla_{\bar{g}^{(2)}} \mathcal{L}) \odot \bar{g}^{(2)} \odot (1 - \bar{g}^{(2)})$
activations$^{(2)}$ $g^{(2)}_i = \sigma(f^{(2)}_i)$	$\bar{g}^{(2)} = \sigma(\bar{f}^{(2)})$	$\frac{\partial \mathcal{L}}{\partial g^{(2)}_i} = \frac{2}{N^{(2)}} (g_i - y_i)$	$\nabla_{\bar{g}^{(2)}} \mathcal{L} = \frac{2}{N^{(2)}} \operatorname{sum}(\bar{g}^{(2)} - \bar{y})$
loss (cost, error) $\mathcal{L} = \frac{1}{N^{(2)}} \sum_i (g^{(2)}_i - y_i)^2$	$\mathcal{L} \in \mathbb{R}$ $\mathcal{L} = \lVert \bar{g}^{(2)} - \bar{y} \rVert^2_2$		