



inputs  
 $x_0, \dots, x_{N^{(0)}-1}$

pre-activations  
 $f_i = \sum_j W_{i,j} x_j + b_i$

activations  
 $g_i = \sigma(f_i)$

pre-activations<sup>(2)</sup>  
 $f_i^{(2)} = \sum_j W_{i,j}^{(2)} g_j + b_i^{(2)}$

activations<sup>(2)</sup>  
 $g_i^{(2)} = \sigma(f_i^{(2)})$

loss (cost, error)  
 $\mathcal{L} = \frac{1}{N^{(2)}} \sum_i (g_i^{(2)} - y_i)^2$

$\mathcal{L}^{(\text{minibatch})} = \frac{1}{|\mathcal{B}|} \sum_{(\bar{x}, \bar{y}) \in \mathcal{B}} \mathcal{L}(\bar{x}, \bar{y})$

forward (as vectors)

$\bar{x} \in \mathbb{R}^{N^{(0)}}$

$\bar{W}^{(1)} \in \mathbb{R}^{N^{(1)} \times N^{(0)}}$   
 $\bar{b}^{(1)} \in \mathbb{R}^{N^{(1)}}$

$\bar{f}^{(1)} \in \mathbb{R}^{N^{(1)}}$   
 $\bar{f}^{(1)} = W^{(1)} \bar{x} + \bar{b}^{(1)}$

$\bar{g}^{(1)} \in \mathbb{R}^{N^{(1)}}$   
 $\bar{g}^{(1)} = \sigma(\bar{f}^{(1)})$

$\bar{f}^{(2)} = W^{(2)} \bar{g}^{(1)} + \bar{b}^{(2)}$

$\bar{g}^{(2)} = \sigma(\bar{f}^{(2)})$

$\mathcal{L} \in \mathbb{R}$

$\mathcal{L} = \|\bar{g}^{(2)} - \bar{y}\|_2^2$

backward pass

$\frac{\partial \mathcal{L}}{\partial b_i} = \frac{\partial \mathcal{L}}{\partial f_i}$

$\frac{\partial \mathcal{L}}{\partial W_{i,j}} = \frac{\partial \mathcal{L}}{\partial f_i} \cdot x_j$

$\frac{\partial \mathcal{L}}{\partial g_j^{(1)}} = \sum_i \frac{\partial \mathcal{L}}{\partial f_i^{(2)}} \cdot W_{i,j}^{(2)}$

$\frac{\partial \mathcal{L}}{\partial f_i^{(2)}} = \frac{\partial \mathcal{L}}{\partial g_i^{(2)}} \cdot \sigma'(f_i^{(2)})$

where  $\sigma' = \sigma \cdot (1 - \sigma)$

$\frac{\partial \mathcal{L}}{\partial g_i^{(2)}} = \frac{2}{N^{(2)}} (g_i - y_i)$

backward (as vectors)

$\nabla_{\bar{b}} \mathcal{L} = \nabla_{\bar{f}^{(1)}} \mathcal{L} \in \mathbb{R}^{N^{(1)}}$

$\nabla_W \mathcal{L} = (\nabla_{\bar{f}^{(1)}} \mathcal{L}) \bar{x}^\top \in \mathbb{R}^{N^{(1)} \times N^{(0)}}$

$\nabla_{\bar{f}^{(1)}} \mathcal{L} = (\nabla_{\bar{g}^{(1)}} \mathcal{L}) \odot \bar{g}^{(1)} \odot (1 - \bar{g}^{(1)})$

$\nabla_{\bar{g}^{(1)}} \mathcal{L} = W^{(2)\top} (\nabla_{\bar{f}^{(2)}} \mathcal{L})$

$\nabla_{\bar{f}^{(2)}} \mathcal{L} = (\nabla_{\bar{g}^{(2)}} \mathcal{L}) \odot \bar{g}^{(2)} \odot (1 - \bar{g}^{(2)})$

$\nabla_{\bar{g}^{(2)}} \mathcal{L} = \frac{2}{N^{(2)}} \text{sum}(\bar{g}^{(2)} - \bar{y})$

$\nabla_{\bar{?}} \mathcal{L}^{(\text{minibatch})} = \frac{1}{|\mathcal{B}|} \text{sum}_{\mathcal{B}} \nabla_{\bar{?}} \mathcal{L}$