$$\mathbf{f}(\mathbf{x}) = \mathbf{W}_2 \cdot h\left(\mathbf{W}_1 \cdot \mathbf{x} + \mathbf{b}_1\right) + \mathbf{b}_2; \qquad \mathbf{f}(\mathbf{x}) \in \mathbb{R}^C, \ \widehat{p}(y = c|\mathbf{x}) = \exp(f_c(\mathbf{x})) / \sum_j \exp(f_j(\mathbf{x}))$$

$$L(\mathbf{x}, y) = -\log \widehat{p}(y|\mathbf{x}) = -f_y(\mathbf{x}) + \log \sum_c \exp\left(f_c(\mathbf{x})\right)$$

$$= -\mathbf{w}_{2,y} \cdot h(\mathbf{W}_1 \cdot \mathbf{x} + \mathbf{b}_1) - b_{2,y} + \log \sum_c \exp\left(\mathbf{w}_{2,c} \cdot h(\mathbf{W}_1 \cdot \mathbf{x} + \mathbf{b}_1) + b_{2,c}\right)$$