预测值为

$$\hat{Y} = g(g(XW^{(1)} + b^{(1)})W^{(2)} + b^{(2)})$$

重新定义：

$$\begin{cases} I = X \\ h_1 = IW^{(1)} + b^{(1)} \\ O_1 = g(h_1) \\ h_2 = O_1W^{(2)} + b^{(2)} \\ O_2 = g(h_2) = \hat{Y} \end{cases} \tag{1}$$

loss function

$$J(W) = \frac{1}{2M}\sum_{i=1}^{M}||Y_i - \hat{Y}_i||^2 = \frac{1}{2M}||Y - O_2||_F^2$$

求反向传播，先考虑 $b^{(2)}$

$$\frac{\partial J}{\partial b^{(2)}} = \frac{1}{M}\sum_{i=1}^{M}(Y_i - O_{2i}) \odot g'(h_{2i})$$

考虑 $W^{(2)}$

$$\frac{\partial J}{\partial W^{(2)}} = O_1^T \cdot ((Y - O_2) \odot g'(h_2))$$

考虑 $b^{(1)}$

$$\frac{\partial J}{\partial b^{(1)}} = \sum_{i=1}^{M}(Y_i - O_{2i}) \cdot W^{(2)T} \odot g'(h_1)$$

考虑 $W^{(1)}$

$$\frac{\partial J}{\partial W^{(1)}} = I^T \cdot (((Y - O_2) \cdot W^{(2)T}) \odot g'(h_1))$$