

$$z^{(2)} = XW^{(1)} \quad (1)$$

$$a^{(2)} = f(z^{(2)}) \quad (2)$$

$$z^{(3)} = a^{(2)}W^{(2)} \quad (3)$$

$$\hat{y} = f(z^{(3)}) \quad (4)$$

$$J = \sum \frac{1}{2}(y - \hat{y})^2 \quad (5)$$

Or

$$J = \sum \frac{1}{2}(y - f(f(XW^{(1)})W^{(2)}))^2 \quad (5.1)$$

$$\frac{\partial J}{\partial W^{(2)}} = (a^{(2)})^T \delta^{(3)} \quad (6.1)$$

$$\delta^{(3)} = -(y - \hat{y})f'(z^{(3)}) \quad (6.2)$$