三层神经网络：一个输入层，一个隐藏层，一个输出层

1. 从输入层到隐藏层

$x^{(1)}$

$$x^{(1)} = \begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \dots \\ x_j^{(1)} \\ \dots \\ x_{n1}^{(1)} \end{pmatrix}_{n1\times 1}$$

X

$$X = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(i)} & \dots & x_1^{(m)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(i)} & \dots & x_2^{(m)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_j^{(1)} & x_j^{(2)} & \dots & x_j^{(i)} & \dots & x_j^{(m)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1}^{(1)} & x_{n1}^{(2)} & \dots & x_{n1}^{(i)} & \dots & x_{n1}^{(m)} \end{pmatrix}_{n1\times m}$$

$y^{(1)}$

$y^{(1)} = y^{(1)}$

Y

$Y = \begin{pmatrix} y^{(1)} & y^{(2)} & \dots & y^{(i)} & \dots & y^{(m)} \end{pmatrix}_{1\times m}$

$W_1^{(1)}$

$$W_1^{(1)} = \begin{pmatrix} W_{1,1}^{(1)} \\ W_{1,2}^{(1)} \\ \dots \\ W_{1,n1}^{(1)} \end{pmatrix}_{n1\times 1}$$

$W^{(1)}$

$$W^{(1)} = \begin{pmatrix} W_{1,1}^{(1)} & W_{2,1}^{(1)} & \dots & W_{j,1}^{(1)} & \dots & W_{n2,1}^{(1)} \\ W_{1,2}^{(1)} & W_{2,2}^{(1)} & \dots & W_{j,2}^{(1)} & \dots & W_{n2,2}^{(1)} \\ \dots \\ W_{1,i}^{(1)} & W_{2,i}^{(1)} & \dots & W_{j,i}^{(1)} & \dots & W_{n2,i}^{(1)} \\ \dots \\ W_{1,n1}^{(1)} & W_{2,n1}^{(1)} & \dots & W_{j,n1}^{(1)} & \dots & W_{n2,n1}^{(1)} \end{pmatrix}_{n1 \times n2}$$

$b_1^{(1)}$

$b^{(1)}$

$$b^{(1)} = \begin{pmatrix} b_1^{(1)} \\ b_2^{(1)} \\ \dots \\ b_{n2}^{(1)} \end{pmatrix}_{n2 \times 1}$$

$z_{1,i}^{(1)}$

$z_{1,i}^{(1)} = (W^{(1)})^T x^{(i)} + b^{(1)} = W_{1,1}^{(1)} x_1^{(i)} + W_{1,2}^{(1)} x_2^{(i)} + \dots + W_{1,n1}^{(1)} x_{n1}^{(i)} + b_1^{(1)}$

$$z_{,i}^{(1)} = \begin{pmatrix} W_{1,1}^{(1)} x_1^{(1)} + W_{1,2}^{(1)} x_2^{(1)} + \dots + W_{1,n1}^{(1)} x_{n1}^{(1)} + b_1^{(1)} \\ W_{2,1}^{(1)} x_1^{(1)} + W_{2,2}^{(1)} x_2^{(1)} + \dots + W_{2,n1}^{(1)} x_{n1}^{(1)} + b_2^{(1)} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ W_{n2,1}^{(1)} x_1^{(1)} + W_{n2,2}^{(1)} x_2^{(1)} + \dots + W_{n2,n1}^{(1)} x_{n1}^{(1)} + b_{n2}^{(1)} \end{pmatrix}_{n2 \times 1}$$

所有样本的第一层$Z^{(1)}$

$$Z^{(1)} = \begin{pmatrix} z_{,1}^{(1)} & z_{,2}^{(1)} & \dots & z_{,m}^{(1)} \end{pmatrix}_{n2 \times m}$$

$$Z^{(1)} = \begin{pmatrix} z_{1,1}^{(1)} & z_{1,2}^{(1)} & \dots & z_{1,m}^{(1)} \\ z_{2,1}^{(1)} & z_{2,2}^{(1)} & \dots & z_{2,m}^{(1)} \\ \dots \\ z_{n2,1}^{(1)} & z_{n2,2}^{(1)} & \dots & z_{n2,m}^{(1)} \end{pmatrix}_{n2 \times m}$$

$a_{1,i}^{(1)}$

$$a_{1,i}^{(1)} = sigmoid(z_{1,i}^{(1)}) = sigmoid((W^{(1)})^T x^{(i)} + b_1^{(1)})$$

所有样本的第一层$A^{(1)}$

$$A^{(1)} = \begin{pmatrix} a_{,1}^{(1)} & a_{,2}^{(1)} & \dots & a_{,m}^{(1)} \end{pmatrix}_{n2 \times m}$$

$$A^{(1)} = \begin{pmatrix} a_{1,1}^{(1)} & a_{1,2}^{(1)} & \dots & a_{1,m}^{(1)} \\ a_{2,1}^{(1)} & a_{2,2}^{(1)} & \dots & a_{2,m}^{(1)} \\ \dots \\ a_{n2,1}^{(1)} & a_{n2,2}^{(1)} & \dots & a_{n2,m}^{(1)} \end{pmatrix}_{n2 \times m}$$

2. 从隐藏层到输出层

$W_1^{(2)}$

$$W_1^{(2)} = \begin{pmatrix} W_{1,1}^{(2)} \\ W_{1,2}^{(2)} \\ \dots \\ W_{1,n2}^{(2)} \end{pmatrix}_{n2\times 1}$$

$W^{(2)}$

$$W^{(2)} = \begin{pmatrix} W_{1,1}^{(2)} \\ W_{1,2}^{(2)} \\ \dots \\ W_{1,n2}^{(2)} \end{pmatrix}_{n2\times 1}$$

$b_1^{(2)}$

$b^{(2)}$

$$b^{(1)} = \begin{pmatrix} b_1^{(2)} \end{pmatrix}_{1\times 1}$$

$z_1^{(2)}$

$z_1^{(2)} = (W^{(2)})^T a^{(1)} + b^{(2)}$

$$z_1^{(2)} = \begin{pmatrix} W_{1,1}^{(2)} a_1^{(1)} + W_{1,2}^{(2)} a_2^{(1)} + \dots + W_{1,n2}^{(1)} a_{n2}^{(1)} + b_1^{(2)} \end{pmatrix}_{1\times 1}$$

$Z^{(2)}$：所有样本的$z^{(2)}$

$$Z^{(2)} = \begin{pmatrix} z_1^{(2)} & z_2^{(2)} & \dots & z_m^{(2)} \end{pmatrix}_{1\times m}$$

$a_1^{(2)}$

$$a_1^{(2)} = sigmoid(z_1^{(2)}) = sigmoid((W^{(2)})^T a^{(1)})$$

$A^{(2)}$：所有样本的$a^{(2)}$

$$A^{(2)} = \begin{pmatrix} a_1^{(2)} & a_2^{(2)} & \dots & a_m^{(2)} \end{pmatrix}_{1\times m}$$

3. 反向传播

每个样本的误差函数$j^{(1)}$

$$j^{(i)} = -[y^{(i)} log(a_{1,i}^{(2)}) + （1 - y^{(i)}) log(1 - a_{1,i}^{(2)})]$$

所有样本的误差函数

$$J=-\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}log(a_{1,i}^{(2)})+（1-y^{(i)})log(1-a_{1,i}^{(2)})]$$

单个样本求偏导

$$j^{(i)}=-[y^{(i)}log(a_{1,i}^{(2)})+（1-y^{(i)})log(1-a_{1,i}^{(2)})]$$

$$\frac{\partial j^{(i)}}{\partial a_{1,i}^{(2)}}=\frac{a_{1,i}^{(2)}-y^{(i)}}{a_{1,i}^{(2)}(1-a_{1,i}^{(2)})}$$

$$\frac{\partial a_{1,i}^{(2)}}{\partial z_{1,i}^{(2)}}=a_{1,i}^{(2)}(1-a_{1,i}^{(2)})$$

$$\frac{\partial z_{1,i}^{(2)}}{\partial W_{1,1}^{(2)}}=a_{1,i}^{(1)}$$

$$\frac{\partial z_{1,i}^{(2)}}{\partial W_{1,2}^{(2)}}=a_{2,i}^{(1)}$$

$$\frac{\partial z_{1,i}^{(2)}}{\partial W_{1,j}^{(2)}}=a_{j,i}^{(1)}$$

$\dots$

$$\frac{\partial z_{1,i}^{(2)}}{\partial W^{(2)}}=\begin{pmatrix}a_{1,i}^{(1)}\\a_{2,i}^{(1)}\\\dots\\a_{j,i}^{(1)}\\\dots\\a_{n2,i}^{(1)}\end{pmatrix}_{n2\times 1}$$

$\frac{\partial z_{1,i}^{(2)}}{\partial b_{1}^{(2)}}=1$

$$\frac{\partial j^{(i)}}{\partial W^{(2)}}=\frac{\partial j^{(i)}}{\partial a_{1,i}^{(2)}}\frac{\partial a_{1,i}^{(2)}}{\partial z_{1,i}^{(2)}}\frac{\partial z_{1,i}^{(2)}}{\partial W^{(2)}}=a_{1,i}^{(2)}-y^{(i)}\begin{pmatrix}a_{1,i}^{(1)}\\a_{2,i}^{(1)}\\\dots\\a_{j,i}^{(1)}\\\dots\\a_{n2,i}^{(1)}\end{pmatrix}_{n2\times 1}$$

$$\frac{\partial j^{(i)}}{\partial W_{1,1}^{(2)}}=\frac{\partial j^{(i)}}{\partial a_{1,i}^{(2)}}\frac{\partial a_{1,i}^{(2)}}{\partial z_{1,i}^{(2)}}\frac{\partial z_{1,i}^{(2)}}{\partial W_{1,1}^{(2)}}=(a_{1,i}^{(2)}-y^{(i)})a_{1,i}^{(1)}$$

$$\frac{\partial j^{(i)}}{\partial W_{1,j}^{(2)}}=\frac{\partial j^{(i)}}{\partial a_{1,i}^{(2)}}\frac{\partial a_{1,i}^{(2)}}{\partial z_{1,i}^{(2)}}\frac{\partial z_{1,i}^{(2)}}{\partial W_{1,j}^{(2)}}=(a_{1,i}^{(2)}-y^{(i)})a_{j,i}^{(1)}$$

$$\frac{\partial J}{\partial W_{1,j}^{(2)}} = \frac{1}{m}\sum_{i=1}^{m}(a_{1,i}^{(2)} - y^{(i)})a_{j,i}^{(1)}$$

$$= \frac{1}{m}[(a_{1,1}^{(2)} - y^{(1)})a_{j,1}^{(1)} + (a_{1,2}^{(2)} - y^{(2)})a_{j,2}^{(1)} + \ldots + (a_{1,m}^{(2)} - y^{(m)})a_{j,m}^{(1)}]$$

$$= \frac{1}{m}\begin{pmatrix} a_{j,1}^{(1)} & a_{j,2}^{(1)} & \ldots & a_{j,m}^{(1)} \end{pmatrix}\begin{pmatrix} a_{1,1}^{(2)} - y^{(1)} \\ a_{1,2}^{(2)} - y^{(2)} \\ \ldots \\ a_{1,m}^{(2)} - y^{(m)} \end{pmatrix}$$

$$\frac{\partial J}{\partial W^{(2)}} = \begin{pmatrix} \frac{\partial J}{\partial W_{1,1}^{(2)}} \\ \frac{\partial J}{\partial W_{1,2}^{(2)}} \\ \ldots \\ \frac{\partial J}{\partial W_{1,j}^{(2)}} \\ \ldots \\ \frac{\partial J}{\partial W_{1,n2}^{(2)}} \end{pmatrix}_{n2\times 1}$$

$$\frac{\partial J}{\partial W^{(2)}} = \frac{1}{m}\begin{pmatrix} a_{1,1}^{(1)} & a_{1,2}^{(1)} & \ldots & a_{1,m}^{(1)} \\ a_{2,1}^{(1)} & a_{2,2}^{(1)} & \ldots & a_{2,m}^{(1)} \\ \ldots \\ a_{j,1}^{(1)} & a_{j,2}^{(1)} & \ldots & a_{j,m}^{(1)} \\ \ldots \\ a_{n2,1}^{(1)} & a_{n2,2}^{(1)} & \ldots & a_{n2,m}^{(1)} \end{pmatrix}_{n2\times m}\begin{pmatrix} a_{1,1}^{(2)} - y^{(1)} \\ a_{1,2}^{(2)} - y^{(2)} \\ \ldots \\ a_{1,m}^{(2)} - y^{(m)} \end{pmatrix}_{m\times 1}$$

$\frac{\partial J}{\partial W^{(2)}}$

$$\frac{\partial J}{\partial W^{(2)}} = \frac{1}{m}A^{(1)}(A^{(2)} - Y)^T$$

$$\frac{\partial j^{(i)}}{\partial b_1^{(2)}} = \frac{\partial j^{(i)}}{\partial a_{1,i}^{(2)}}\frac{\partial a_{1,i}^{(2)}}{\partial z_{1,i}^{(2)}}\frac{\partial z_{1,i}^{(2)}}{\partial b_1^{(2)}} = (a_{1,i}^{(2)} - y^{(i)})$$

$\frac{\partial J}{\partial b^{(2)}}$

$$\frac{\partial J}{\partial b^{(2)}} = \frac{1}{m}\sum_{i=1}^{m}\frac{\partial j^{(i)}}{\partial b_1^{(2)}} = \frac{1}{m}\sum_{i=1}^{m}(a_{1,i}^{(2)} - y^{(i)}) = \frac{1}{m}\sum_{i=1}^{m}dZ^{(2)} = \frac{1}{m}\sum_{i=1}^{m}(A^{(2)} - Y) = \frac{1}{m}np.sum(A^{(2)} - Y)$$

从隐藏层像输入层传播

$$\frac{\partial z_{1,i}^{(2)}}{\partial a_{j,i}^{(1)}} = W_{1,j}^{(2)}$$

$$\frac{\partial a_{j,i}^{(1)}}{\partial z_{j,i}^{(1)}} = a_{j,i}^{(1)}(1 - a_{j,i}^{(1)})$$

$$\frac{\partial z_{j,i}^{(1)}}{\partial W_{j,k}^{(1)}} = x_k^{(i)}$$

$$\frac{\partial z_{j,i}^{(1)}}{\partial b_j^{(1)}} = 1$$

$$\frac{\partial j^{(i)}}{\partial W_{j,k}^{(1)}} = \frac{\partial j^{(i)}}{\partial a_{1,i}^{(2)}} \frac{\partial a_{1,i}^{(2)}}{\partial z_{1,i}^{(2)}} \frac{\partial z_{1,i}^{(2)}}{\partial a_{j,i}^{(1)}} \frac{\partial a_{j,i}^{(1)}}{\partial z_{j,i}^{(1)}} \frac{\partial z_{j,i}^{(1)}}{\partial W_{j,k}^{(1)}} = \frac{a_{1,i}^{(2)} - y^{(i)}}{a_{1,i}^{(2)}(1 - a_{1,i}^{(2)})} * a_{1,i}^{(2)}(1 - a_{1,i}^{(2)}) * W_{1,j}^{(2)} * a_{j,i}^{(1)}(1 - a_{j,i}^{(1)}) * x_k^{(i)}$$

$$\frac{\partial j^{(i)}}{\partial b_j^{(1)}} = \frac{\partial j^{(i)}}{\partial a_{1,i}^{(2)}} \frac{\partial a_{1,i}^{(2)}}{\partial z_{1,i}^{(2)}} \frac{\partial z_{1,i}^{(2)}}{\partial a_{j,i}^{(1)}} \frac{\partial a_{j,i}^{(1)}}{\partial z_{j,i}^{(1)}} \frac{\partial z_{j,i}^{(1)}}{\partial b_j^{(1)}} = \frac{a_{1,i}^{(2)} - y^{(i)}}{a_{1,i}^{(2)}(1 - a_{1,i}^{(2)})} * a_{1,i}^{(2)}(1 - a_{1,i}^{(2)}) * W_{1,j}^{(2)} * a_{j,i}^{(1)}(1 - a_{j,i}^{(1)})$$

$$\frac{\partial j^{(i)}}{\partial W_{j,k}^{(1)}} = \frac{\partial j^{(i)}}{\partial a_{1,i}^{(2)}} \frac{\partial a_{1,i}^{(2)}}{\partial z_{1,i}^{(2)}} * [\frac{\partial z_{1,i}^{(2)}}{\partial W_{j,k}^{(1)}}]$$

$$\frac{\partial z_{1,i}^{(2)}}{\partial W_{j,k}^{(1)}} = \sum_{j=1}^{n2} \frac{\partial z_{1,i}^{(2)}}{\partial a_{j,i}^{(1)}} \frac{\partial a_{j,i}^{(1)}}{\partial z_{j,i}^{(1)}} \frac{\partial z_{j,i}^{(1)}}{\partial W_{j,k}^{(1)}} = \sum_{j=1}^{n2} W_{1,j}^{(2)} * a_{j,i}^{(1)}(1 - a_{j,i}^{(1)}) * x_k^{(i)}$$

$$\frac{\partial z_{1,i}^{(2)}}{\partial a_{j,i}^{(1)}} \frac{\partial a_{j,i}^{(1)}}{\partial z_{j,i}^{(1)}} \frac{\partial z_{j,i}^{(1)}}{\partial W_{j,k}^{(1)}} = \frac{a_{1,i}^{(2)} - y^{(i)}}{a_{1,i}^{(2)}(1 - a_{1,i}^{(2)})} * a_{1,i}^{(2)}(1 - a_{1,i}^{(2)}) * W_{1,j}^{(2)} * a_{j,i}^{(1)}(1 - a_{j,i}^{(1)}) * x_k^{(i)}$$

$$\frac{\partial j^{(i)}}{\partial W_{j,k}^{(1)}} = (a_{1,i}^{(2)} - y^{(i)}) * W_{1,j}^{(2)} * a_{j,i}^{(1)}(1 - a_{j,i}^{(1)}) * x_k^{(i)}$$

$$\frac{\partial j^{(1)}}{\partial W_{j,k}^{(1)}} = (a_{1,1}^{(2)} - y^{(1)}) * W_{1,j}^{(2)} * a_{j,1}^{(1)}(1 - a_{j,1}^{(1)}) * x_k^{(1)}$$

$$\frac{\partial j^{(1)}}{\partial W_{1,1}^{(1)}} = (a_{1,1}^{(2)} - y^{(1)}) * W_{1,1}^{(2)} * a_{1,1}^{(1)}(1 - a_{1,1}^{(1)}) * x_1^{(1)}$$

$$\frac{\partial j^{(1)}}{\partial W_{1,2}^{(1)}} = (a_{1,1}^{(2)} - y^{(1)}) * W_{1,1}^{(2)} * a_{1,1}^{(1)}(1 - a_{1,1}^{(1)}) * x_2^{(1)}$$

$$\frac{\partial j^{(1)}}{\partial W_{j,k}^{(1)}} = (a_{1,1}^{(2)} - y^{(1)}) * W_{1,1}^{(2)} * a_{1,1}^{(1)}(1 - a_{1,1}^{(1)}) * x_3^{(1)}$$

$$\frac{\partial j^{(1)}}{\partial W_{j,k}^{(1)}} = (a_{1,1}^{(2)} - y^{(1)}) * W_{1,1}^{(2)} * a_{1,1}^{(1)}(1 - a_{1,1}^{(1)}) * x_4^{(1)}$$

$$\frac{\partial J}{\partial W_{j,k}^{(1)}}$$

$$\frac{\partial J}{\partial W_{j,k}^{(1)}} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial j^{(i)}}{\partial W_{j,k}^{(1)}} = \frac{1}{m} \sum_{i=1}^{m} [(a_{1,i}^{(2)} - y^{(i)}) * W_{1,j}^{(2)} * a_{j,i}^{(1)}(1 - a_{j,i}^{(1)}) * x_k^{(i)}]$$

$$\frac{\partial J}{\partial b_j^{(1)}} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial j^{(i)}}{\partial W_{j,k}^{(1)}} = \frac{1}{m} \sum_{i=1}^{m} [(a_{1,i}^{(2)} - y^{(i)}) * W_{1,j}^{(2)} * a_{j,i}^{(1)}(1 - a_{j,i}^{(1)})]$$

$$\frac{\partial J}{\partial W^{(1)}} = \begin{pmatrix} \frac{\partial J}{\partial b_1^{(1)}} \\ \frac{\partial J}{\partial b_2^{(1)}} \\ \dots \\ \frac{\partial J}{\partial b_j^{(1)}} \\ \dots \\ \frac{\partial J}{\partial b_{n2}^{(1)}} \end{pmatrix}_{n2\times 1}$$

$$\frac{\partial J}{\partial W^{(1)}} = \begin{pmatrix} \frac{\partial J}{\partial W_{1,1}^{(1)}} & \frac{\partial J}{\partial W_{2,1}^{(1)}} & \dots & \frac{\partial J}{\partial W_{j,1}^{(1)}} & \dots & \frac{\partial J}{\partial W_{n2,1}^{(1)}} \\ \frac{\partial J}{\partial W_{1,2}^{(1)}} & \frac{\partial J}{\partial W_{2,2}^{(1)}} & \dots & \frac{\partial J}{\partial W_{j,2}^{(1)}} & \dots & \frac{\partial J}{\partial W_{n2,2}^{(1)}} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial J}{\partial W_{1,k}^{(1)}} & \frac{\partial J}{\partial W_{2,k}^{(1)}} & \dots & \frac{\partial J}{\partial W_{j,k}^{(1)}} & \dots & \frac{\partial J}{\partial W_{n2,k}^{(1)}} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial J}{\partial W_{1,n1}^{(1)}} & \frac{\partial J}{\partial W_{2,n1}^{(1)}} & \dots & \frac{\partial J}{\partial W_{j,n1}^{(1)}} & \dots & \frac{\partial J}{\partial W_{n2,n1}^{(1)}} \end{pmatrix}_{n1\times n2}$$

$$\begin{pmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(i)} & \dots & x_1^{(m)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(i)} & \dots & x_2^{(m)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_j^{(1)} & x_j^{(2)} & \dots & x_j^{(i)} & \dots & x_j^{(m)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1}^{(1)} & x_{n1}^{(2)} & \dots & x_{n1}^{(i)} & \dots & x_{n1}^{(m)} \end{pmatrix}_{n1\times m} \begin{pmatrix} a_{1,1}^{(2)} - y^{(1)} \\ a_{1,2}^{(2)} - y^{(2)} \\ \dots \\ a_{1,m}^{(2)} - y^{(m)} \end{pmatrix}_{m\times 1} \odot \begin{pmatrix} a_{1,1}^{(1)}(1-a_{1,1}^{(1)}) & a_{2,1}^{(1)}(1-a_{2,1}^{(1)}) & \dots & a_{n2,1}^{(1)}(1-a_{n2,1}^{(1)}) \\ a_{1,2}^{(1)}(1-a_{1,2}^{(1)}) & a_{2,2}^{(1)}(1-a_{2,2}^{(1)}) & \dots & a_{n2,2}^{(1)}(1-a_{n2,2}^{(1)}) \\ \dots \\ a_{1,m}^{(1)}(1-a_{1,m}^{(1)}) & a_{2,m}^{(1)}(1-a_{2,m}^{(1)}) & \dots & a_{n2,m}^{(1)}(1-a_{n2,m}^{(1)}) \end{pmatrix}_{m\times n2} \odot \begin{pmatrix} W_{1,1}^{(2)} & W_{1,2}^{(2)} & \dots & W_{1,n2}^{(2)} \end{pmatrix}_{1\times n2}$$

$$(A^{(2)} - Y)^T (W^{(2)})^T$$

$$\begin{pmatrix} a_{1,1}^{(2)} - y^{(1)} \\ a_{1,2}^{(2)} - y^{(2)} \\ \dots \\ a_{1,m}^{(2)} - y^{(m)} \end{pmatrix}_{m\times 1} \begin{pmatrix} W_{1,1}^{(2)} & W_{1,2}^{(2)} & \dots & W_{1,n2}^{(2)} \end{pmatrix}_{1\times n2} = \begin{pmatrix} (a_{1,1}^{(2)} - y^{(1)})W_{1,1}^{(2)} & (a_{1,1}^{(2)} - y^{(1)})W_{1,2}^{(2)} & \dots (a_{1,1}^{(2)} - y^{(1)})W_{1,n2}^{(2)} \\ (a_{1,2}^{(2)} - y^{(2)})W_{1,1}^{(2)} & (a_{1,2}^{(2)} - y^{(2)})W_{1,2}^{(2)} & \dots (a_{1,2}^{(2)} - y^{(2)})W_{1,n2}^{(2)} \\ \dots & \dots & \dots \\ (a_{1,m}^{(2)} - y^{(m)})W_{1,1}^{(2)} & (a_m^{(2)} - y^{(m)})W_{1,2}^{(2)} & \dots (a_{1,m}^{(2)} - y^{(m)})W_{1,n2}^{(2)} \end{pmatrix}_{m\times n2}$$

$$[(A^{(2)} - Y)^T (W^{(2)})^T] \odot (\sigma' Z^{(2)})^T$$

$$\begin{pmatrix} (a_{1,1}^{(2)} - y^{(1)})W_{1,1}^{(2)} & (a_{1,1}^{(2)} - y^{(1)})W_{1,2}^{(2)} & \dots & (a_{1,1}^{(2)} - y^{(1)})W_{1,n2}^{(2)} \\ (a_{1,2}^{(2)} - y^{(2)})W_{1,1}^{(2)} & (a_{1,2}^{(2)} - y^{(2)})W_{1,2}^{(2)} & \dots & (a_{1,2}^{(2)} - y^{(2)})W_{1,n2}^{(2)} \\ \dots & \dots & \dots \\ (a_{1,m}^{(2)} - y^{(m)})W_{1,1}^{(2)} & (a_m^{(2)} - y^{(m)})W_{1,2}^{(2)} & \dots & (a_{1,m}^{(2)} - y^{(m)})W_{1,n2}^{(2)} \end{pmatrix}_{m\times n2} \odot \begin{pmatrix} a_{1,1}^{(1)}(1-a_{1,1}^{(1)}) & a_{2,1}^{(1)}(1-a_{2,1}^{(1)}) & \dots & a_{n2,1}^{(1)}(1-a_{n2,1}^{(1)}) \\ a_{1,2}^{(1)}(1-a_{1,2}^{(1)}) & a_{2,2}^{(1)}(1-a_{2,2}^{(1)}) & \dots & a_{n2,2}^{(1)}(1-a_{n2,2}^{(1)}) \\ \dots \\ a_{1,m}^{(1)}(1-a_{1,m}^{(1)}) & a_{2,m}^{(1)}(1-a_{2,m}^{(1)}) & \dots & a_{n2,m}^{(1)}(1-a_{n2,m}^{(1)}) \end{pmatrix}_{m\times n2}$$

$$\frac{\partial J}{\partial W^{(1)}} = X\{[(A^{(2)} - Y)^T (W^{(2)})^T] \odot (\sigma' Z^{(2)})^T\}$$

$$\frac{\partial J}{\partial b^{(1)}} = [(A^{(2)} - Y)^T (W^{(2)})^T] \odot (\sigma' Z^{(2)})^T$$

$$\begin{pmatrix} a_{1,1}^{(1)}(1-a_{1,1}^{(1)}) & a_{1,2}^{(1)}(1-a_{1,2}^{(1)}) & \dots & a_{1,m}^{(1)}(1-a_{1,m}^{(1)}) \\ a_{2,1}^{(1)}(1-a_{2,1}^{(1)}) & a_{2,2}^{(1)}(1-a_{2,2}^{(1)}) & \dots & a_{2,m}^{(1)}(1-a_{2,m}^{(1)}) \\ a_{n2,1}^{(1)}(1-a_{n2,1}^{(1)}) & a_{n2,2}^{(1)}(1-a_{n2,2}^{(1)}) & \dots & a_{n2,m}^{(1)}(1-a_{n2,m}^{(1)}) \end{pmatrix}_{n2\times m}$$

$$A^{(2)} - Y$$

$$\begin{pmatrix} a_{1,1}^{(2)} - y^{(1)} & a_{1,2}^{(2)} - y^{(2)} & \dots & a_{1,m}^{(2)} - y^{(m)} \end{pmatrix}_{1\times m}$$

$$\frac{\partial j^{(2)}}{\partial W_{j,k}^{(1)}} = (a_{1,2}^{(2)} - y^{(2)}) * W_{1,j}^{(2)} * a_{j,2}^{(1)}(1 - a_{j,2}^{(1)}) * x_k^{(2)}$$

Forward Propagation:

- You get X
- You compute $A = \sigma(w^T X + b) = (a^{(0)}, a^{(1)}, \ldots, a^{(m-1)}, a^{(m)})$
- You calculate the cost function: $J = -\frac{1}{m} \sum_{i=1}^{m} y^{(i)} \log(a^{(i)}) + (1 - y^{(i)}) \log(1 - a^{(i)})$

$$\begin{matrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & z & z^2 \end{matrix}$$

$$\begin{bmatrix} a_{00} & a_{01} \\ \\ a_{10} & a_{11} \end{bmatrix}$$

$$\begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_j^{(1)} \\ x_{n1}^{(1)} \end{bmatrix}$$