back_prop

1、为方便推导，将W1、W2等矩阵拆分，先假设几个参数

输入层、隐层、输出层分别为第1、2、3层

Wi_{jk}表示第i层第k个神经元到i层第j个神经元的权重

bi_j表示第i层第j个神经元的偏置

zi_j表示第i层第j个神经元的输入

$zi_j = \sum_k Wi_{jk} a(i-1)_k + bi_j$

ai_j表示第i层第j个神经元的输出

$ai_j = ReLU(\sum_k Wi_{jk} a(i-1)_k + bi_j)$

2、代价函数

$L = CrossEntrophy(Softmax(a3), y) + \frac{\alpha}{2}\sum ||W||_2^2$

a4表示经过Softmax层后的输出

$a4_j = \frac{e^{a3_j}}{\sum_n e^{a3_n}}$

3中计算先假设无L2正则项

3、神经网络产生的错误

对于交叉熵与Softmax，任意一个结点的错误

$\delta 4_j = \frac{\partial L}{\partial a3_j} = \sum_k \frac{\partial L_k}{\partial a4_k} \cdot \frac{\partial a4_k}{\partial a3_j}$

$\frac{\partial L_k}{\partial a4_k} = \frac{\partial(-y_k ln a4_k)}{\partial a4_k} = -y_k \frac{1}{a4_k}$

①若j=k

$\frac{\partial a4_j}{\partial a3_j} = \frac{\partial(\frac{e^{a3_j}}{\sum_n e^{a3_n}})}{\partial a3_j} = \frac{\sum_n e^{a3_n} e^{a3_j} - (e^{z_j})^2}{\sum_n e^{a3_n}} = (\frac{e^{a3_j}}{\sum_n e^{a3_n}})(1 - \frac{e^{a3_j}}{\sum_n e^{a3_n}}) = a4_j(1 - a4j)$

②若j不等于k

$\frac{\partial a4_k}{\partial a3_j} = \frac{\partial(\frac{e^{a3_k}}{\sum_n e^{a3_n}})}{\partial a3_j} = -e^{a3_k}(\frac{1}{\sum_n e^{a3_n}})^2 = -a4_j a4_k$

所以

$\delta 4_j = \sum_{j \neq k} -y_k \frac{1}{a4_k}(-a4_j a4_k) + (-y_k \frac{1}{a4_k})(a4_j(1 - a4_j)) = a4_j \sum_k y_k - y_j$

因为给定的分类结果均为一个类别是1，其它为0

所以

$\delta 4_j = a4_j - yj$

对于隐层-输出层，任意一个结点的错误

$\delta 3_j = \frac{\partial L}{\partial z3_j} = \frac{\partial L}{\partial a3_j} \cdot \frac{\partial a3_j}{\partial z3_j} = \delta 4_j ReLU'(z3)$

ReLU'(x) 在x大于0时为1，小于等于0时为0

对于输出层-隐层，任意一个结点的错误

$\delta 2_j = \frac{\partial L}{\partial z2_j} = \sum_k \frac{\partial L}{\partial z3_k} \cdot \frac{\partial z3_k}{\partial a2_j} \cdot \frac{\partial a2_j}{\partial z2_j} = \sum_k \delta 3_k \cdot W2_{kj} \cdot ReLU'(z2_j)$

4、权重W的梯度

对于W2，加入L2正则项后，相当于加了一项 $\lambda W2_j$

$\frac{\partial L}{\partial W2_{jk}} = \frac{\partial L}{\partial z3_j} \cdot \frac{\partial z3_j}{\partial W2_{jk}} = \delta 3_j \cdot \frac{\partial (W2_{jk} \cdot a2_k + b2_j)}{\partial W2_{jk}} = a2_k \delta 3_j + \lambda W2_j = a2_k((a4_j - y_j)ReLU'(z3)) + \lambda W2_j$

对于W1同理有

$\frac{\partial L}{\partial W1_{jk}} = a1_k \delta 2_j + \lambda W1_j$

5、偏置b的梯度

对于b2

$\frac{\partial L}{\partial b2_j} = \frac{\partial L}{\partial z3_j} \cdot \frac{\partial z3_j}{\partial b2_j} = \delta 3_j \cdot \frac{\partial (W2_{jk} \cdot a2_k + b2_j)}{\partial b2_j} = \delta 3_j = (a4_j - y_j)ReLU'(z3)$

对于b1同理有

$\frac{\partial L}{\partial b1_j} = \delta 2_j$