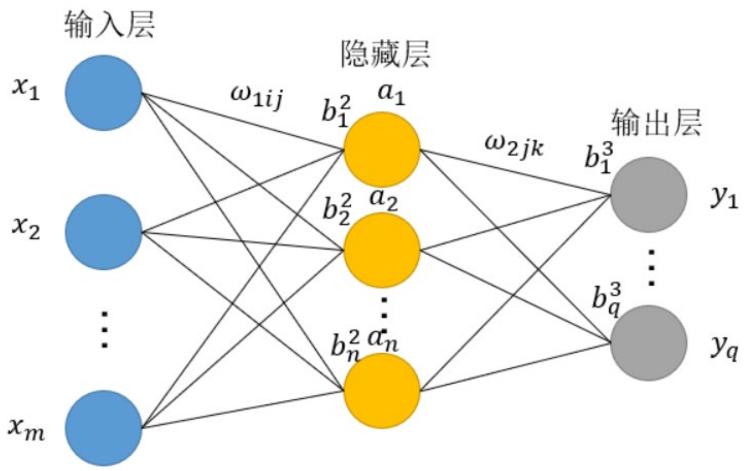
机器学习第三次作业

姓名：刘文晨　　　　　　学号：202222280328

误差反向传播算法是通过误差函数计算实际输出值与期望输出值之间的误差，把误差从最后的输出层依次传播到之前各层，最后通过调整各层连接权重与偏置达到减小误差的目的。而权重和偏置的调整一般使用梯度下降法。

以包含隐藏层的多层感知器为例，结构如下图所示：



图中和推导过程中涉及的符号代表的含义如下表所示：

符号	含义
x_i	输出值
a_j	隐藏层激活值
y_k	实际输出值
r_k	期望输出值
w_{1ij}	网络输入层的第i个神经元和下一层网络第j个神经元之间的连接权重
w_{2jk}	网络隐藏层的第j个神经元和下一层网络第k个神经元之间的连接权重
b_j^2	网络隐藏层第j个神经元的偏置
b_k^3	网络输出层第k个神经元的偏置
u_{1j}	隐藏层第j个神经元的激活函数的加权输入
u_{2k}	输出层第k个神经元的激活函数的加权输入

E	误差函数
η	学习率

误差函数采用最小二乘误差函数，公式如下：

$$E=\sum_{n=1}^{N}\|r_n-y_n\|^2$$

激活函数使用 sigmoid 函数，公式如下：

$$f(u)=\frac{1}{1+e^{-u}}$$

其中，$u=\sum_{i=1}^{N}w_ix_i+b$。

我们的目的是计算使用梯度下降法时神经网络各层间连接权重和偏差的变化值，即Δw_{1ij}，Δw_{2jk}，Δb_j^2等值。以计算Δw_{1ij}为例：

$$\Delta w_{1ij}=\eta\frac{\partial E}{\partial w_{1ij}}$$

由链式法则，求导得：

$$\frac{\partial E}{\partial w_{1ij}}=\sum_{k=1}^{q}\left(\frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial u_{2k}}\frac{\partial u_{2k}}{\partial a_j}\frac{\partial a_j}{\partial u_{1j}}\frac{\partial u_{1j}}{\partial w_{1ij}}\right)$$

然后逐个对等式右侧各部分求解得：

$$\frac{\partial E}{\partial y_k}=-(r_k-y_k)$$

$$\frac{\partial y_k}{\partial u_{2k}}=f(u_{2k})\big(1-f(u_{2k})\big)=y_k(1-y_k)$$

$$\frac{\partial u_{2k}}{\partial a_j}=w_{2jk}$$

$$\frac{\partial a_j}{\partial u_{1j}}=a_j(1-a_j)$$

$$\frac{\partial u_{1j}}{\partial w_{1ij}}=x_i$$

最后得：

$$\frac{\partial E}{\partial w_{1ij}}=-\sum_{k=1}^{q}[(r_k-y_k)y_k(1-y_k)w_{2jk}a_j(1-a_j)x_i]$$

因此，权重w_{1ij}的变化值为：

$$\Delta w_{1ij}=\eta\sum_{k=1}^{q}[(r_k-y_k)y_k(1-y_k)w_{2jk}a_j(1-a_j)x_i]$$

可以发现，权重的变化值的组成依次为：学习率、误差函数导数、激活函数导数、对应连接权重、激活函数导数和输入值，根据这个规律可以快速写出各层权重和偏差的变化值，例如：

$$\Delta w_{2jk}=\eta\sum_{k=1}^{q}[(r_k-y_k)y_k(1-y_k)a_j]$$

$$\Delta b_j^2=\eta\sum_{k=1}^{q}[(r_k-y_k)y_k(1-y_k)w_{2jk}a_j(1-a_j)]$$