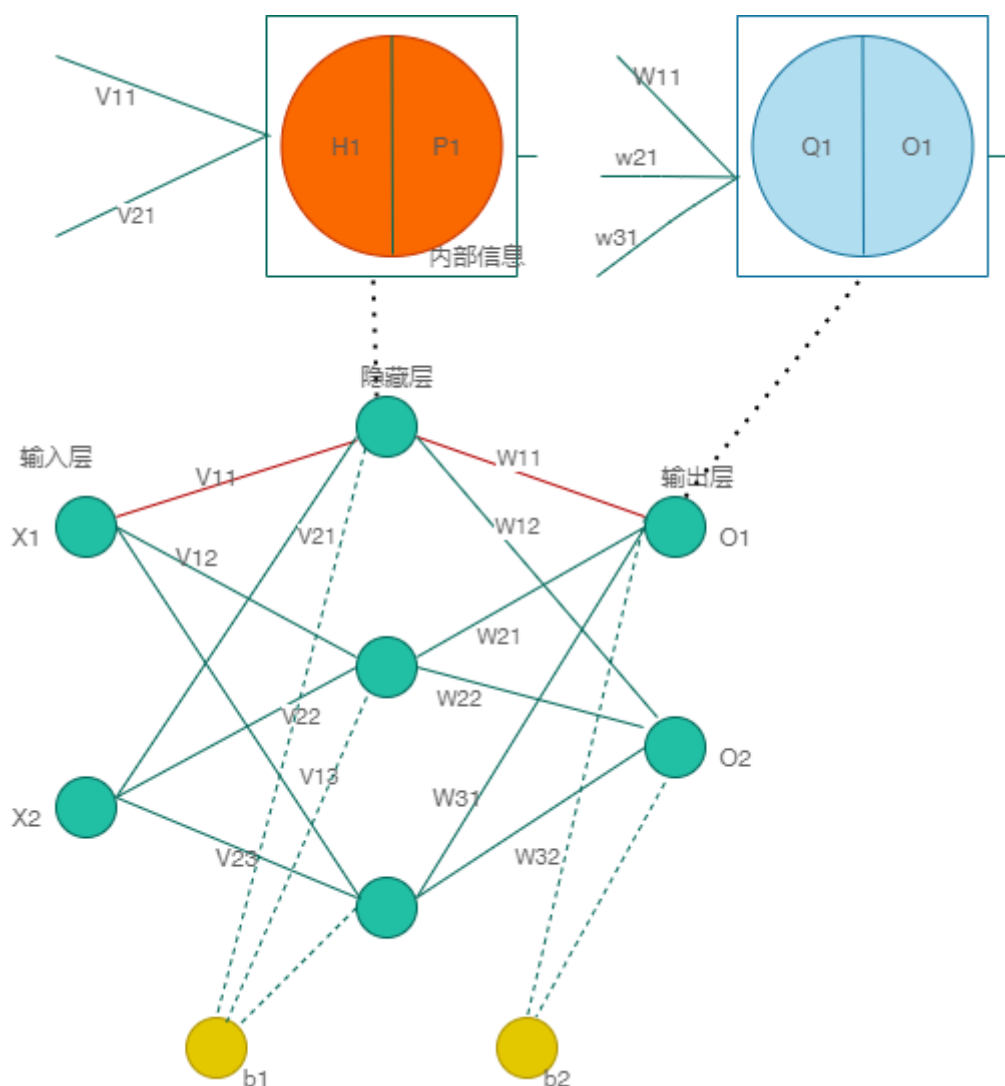


以三层神经网络的正向和反向传播推导



各符号相关说明

如上图描述：

- 1) 输入层，输入 $\{(X1, Y1), (X2, Y2)\}$ ，Y1和Y2为真实的标签
- 2) 隐藏层：我们这里隐藏层有3个神经元，三个神经元的输入分别为：H1，H2，H3，隐藏层的神经元节点输出分别为：P1,P2,P3
- 3) 输出层，我们这里有两个输出，输出层神经元的输入分别为：Q1，Q2，输出层神经元的输出分别为O1，O2.

其中，V11,V12 ... V和W表示权重。Vij表示输入层第i个节点对隐藏层第j个节点的连接权重，

Wij表示隐藏层第i个节点对输出层第j个节点的连接权重。以V11和W11为例进行正向和反向权重更新推导。

为了方便推导，我们假设我们的损失函数为：

$$E = \frac{1}{N} \sum_{i=1}^N (y_i - y_o)^2 \quad (1)$$

同时假设激活函数我们使用sigmoid函数：

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

我们很容易求得sigmoid导数为：

$$f'(x) = f(x)(1 - f(x)) \quad (3)$$

## 正向传播

隐藏层第一个神经元的输入标记为H1，则

$$H1 = V11 * X1 + V21 * X2 + b1 \quad (4)$$

隐藏层第一个神经元的输出标记为P1，也就是H1经过sigmoid后的输出

$$P1 = \text{sigmoid}(H1) = \frac{1}{1 + e^{-H1}} \quad (5)$$

输出层第一个神经元的输入记为Q1，则

$$Q1 = W11 * P1 + W21 * P2 + W31 * P3 + b2 \quad (6)$$

输出层第二个神经元的输入记为Q2，则

$$Q2 = W12 * P1 + W22 * P2 + W32 * P3 + b2 \quad (7)$$

输出层第一个神经元的最终输出记录为O1，则

$$O1 = \text{sigmoid}(Q1) = \frac{1}{1 + e^{-Q1}} \quad (8)$$

输出层第二个神经元的最终输出记录为O2，则

$$O2 = \text{sigmoid}(Q2) = \frac{1}{1 + e^{-Q2}} \quad (9)$$

这样就得到了我们输出层第一个神经元的输出O1,输出层第二个神经元输出O2，完成一次正向传播。

## 反向传播

1.我们首先根据损失函数得到总的误差：

$$E_{total} = \frac{1}{N} \sum_{i=1}^N (y_i - y_o)^2 = \frac{1}{2} ((Y1 - O1)^2 + (Y2 - O2)^2) \quad (10)$$

2.我们先求权重W11的更新，由导数链式法则，我们可以得到

$$\frac{\partial E_{total}}{\partial W11} = \frac{\partial E_{total}}{\partial O1} * \frac{\partial O1}{\partial Q1} * \frac{\partial Q1}{\partial W11} \quad (11)$$

其中：

$$\frac{\partial E_{total}}{\partial O1} = -(Y1 - O1) \quad (12)$$

$$\frac{\partial O1}{\partial Q1} = f(Q1) * (1 - f(Q1)) = \frac{1}{1 + e^{-Q1}} * (1 - \frac{1}{1 + e^{-Q1}}) = O1 * (1 - O1) \quad (13)$$

$$\frac{\partial Q1}{\partial W11} = \frac{\partial (W11 * P1 + W21 * P2 + w31 * P3 + b2)}{\partial W11} = P1 \quad (14)$$

3.更新权重参数W11,由公式 ( 11 ) - ( 14 ) 得：

$$W11 = W11 - \eta * \frac{\partial E_{total}}{\partial W11} = W11 - \eta * (-P1 * (Y1 - O1) * O1 * (1 - O1)) \quad (15)$$

4.接着我们继续求权重V11的参数更新，同样由导数的链式法则，我们得到：

$$\frac{\partial E_{total}}{\partial V11} = \frac{\partial E_{total}}{\partial P1} * \frac{\partial P1}{\partial V11} \quad (16)$$

其中我们知道输出的总误差由两部分数据产生，分别记录为E1和E2。则我们可以得到：

$$\frac{\partial E_{total}}{\partial V11} = \frac{\partial E_{total}}{\partial P1} * \frac{\partial P1}{\partial V11} = (\frac{\partial E1}{\partial P1} + \frac{\partial E2}{\partial P1}) * \frac{\partial P1}{\partial V11} \quad (17)$$

其中

$$E1 = (Y1 - O1)^2 \quad (18)$$

$$E2 = (Y2 - O2)^2 \quad (19)$$

分别求上述公式的拆分项

$$\frac{\partial E1}{\partial P1} = \frac{\partial E1}{\partial O1} * \frac{\partial O1}{\partial Q1} * \frac{\partial Q1}{\partial P1} \quad (20)$$

其中

$$\frac{\partial E1}{\partial O1} = -(Y1 - O1) \quad (21)$$

$$\frac{\partial O1}{\partial Q1} = f(Q1) * (1 - f(Q1)) = \frac{1}{1 + e^{-Q1}} * (1 - \frac{1}{1 + e^{-Q1}}) = O1 * (1 - O1) \quad (22)$$

$$\frac{\partial Q1}{\partial P1} = \frac{\partial (W11 * P1 + W21 * P2 + w31 * P3 + b2)}{\partial P1} = W11 \quad (23)$$

将公式 ( 21 ) -公式 ( 23 ) 代入公式(20)得：

$$\frac{\partial E1}{\partial P1} = -W11 * (Y1 - O1) * O1 * (1 - O1) \quad (24)$$

同理，我们可以求出

$$\frac{\partial E2}{\partial P1} = \frac{\partial E2}{\partial O2} * \frac{\partial O2}{\partial Q2} * \frac{\partial Q2}{\partial P1} \quad (25)$$

其中

$$\frac{\partial E2}{\partial O2} = -(Y2 - O2) \quad (26)$$

$$\frac{\partial O2}{\partial Q2} = f(Q2) * (1 - f(Q2)) = \frac{1}{1 + e^{-Q2}} * (1 - \frac{1}{1 + e^{-Q2}}) = O2 * (1 - O2) \quad (27)$$

$$\frac{\partial Q2}{\partial P1} = \frac{\partial(W12 * P1 + W22 * P2 + w32 * P3 + b2)}{P1} = W12 \quad (28)$$

将公式 ( 26 ) -公式 ( 28 ) 代入公式 ( 25 ) 得：

$$\frac{\partial E2}{\partial P1} = -W12 * (Y2 - O2) * O2 * (1 - O2) \quad (29)$$

最后我们求得：

$$\frac{\partial P1}{\partial V11} = \frac{\partial P1}{\partial H1} * \frac{\partial H1}{\partial V11} \quad (30)$$

其中：

$$\frac{\partial P1}{\partial H1} = f(H1) * (1 - f(H1)) = \frac{1}{1 + e^{-H1}} * (1 - \frac{1}{1 + e^{-H1}}) = P1 * (1 - P1) \quad (31)$$

$$\frac{\partial H1}{\partial V11} = \frac{\partial(V11 * X1 + V21 * X2 + b1)}{\partial V11} = X1 \quad (32)$$

将公式 ( 31 ) 和公式 ( 32 ) 代入公式 ( 30 ) 得：

$$\frac{\partial P1}{\partial V11} = X1 * P1 * (1 - P1) \quad (33)$$

最终V11的梯度更新公式为：

$$V11 = V11 - \eta * \frac{\partial E_{total}}{\partial V11} \quad (34)$$

将公式 ( 24 )、公式 ( 29 ) 和公式 ( 33 ) 代入公式 ( 17 )：

$$\frac{\partial E_{total}}{\partial V11} = (-W11 * (Y1 - O1) * O1 * (1 - O1)) * (-W12 * (Y2 - O2) * O2 * (1 - O2) * (X1 * P1 * (1 - P1))) \quad (35)$$

将公式 ( 35 ) 代入公式 ( 34 )，这样完成了V11权重的更新，至此一次的V11权重更新结束，继续下一次的迭代更新。