多层感知机

2020年5月15日 星期五 22:01

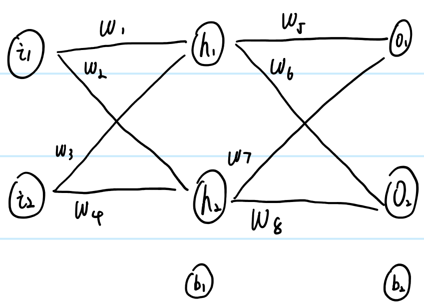
多层感知机与全连接层类似

基本结构就是在每个隐含层后加上激活函数

多层感知机(全连接网络)的灵魂就在于反向传播过程

也是后面所有深度学习的基础，

假设有这样一个网络



注意这里的w.~~感~~前感知机的w不一样.前面感知机的w是一个向量

这里的w是一个数字.前面的输入层有两个神经元.可以看作是输入是一个二维向量.向量中两个值分别进入i_1和i_2

这里b_1和b_2是两个偏置.一般一整个层共用一个偏置

现在假设一个数据集是 $(\underset{i_1}{0.05},\ \underset{i_2}{0.1})-(0.01,\ 0.99)$

设初始权重为 $w_1=0.15$ $w_2=0.2$ $w_3=0.25$ $w_4=0.3$ $b_1=0.35$

$w_5=0.4$ $w_6=0.45$ $w_7=0.5$ $w_8=0.55$ $b_2=0.6$

现在来模拟前向传播过程

(1). 输入层到隐含层

h_1的输入: $h_{1in} = w_1\cdot i_1 + w_2\cdot i_2 + b_1$

$$= 0.15\times 0.05 + 0.2\times 0.1 + 0.35 = 0.3775$$

h_1经激活函数(sigmoid)后的输出:

$$h_{1out} = \frac{1}{1+e^{-h_{1in}}} = \frac{1}{1+e^{-0.3775}} = 0.5932$$

同理算出 $h_{2out} = 0.5968$

(2). 隐含层到输出层

(2). 隐含层到输出层

O_1的输入：$O_{1in} = w_5 \cdot h_{1out} + w_6 \cdot h_{2out} + b_2$

$$= 0.4 \times 0.5932 + 0.45 \times 0.5968 + 0.6 = 1.1059$$

O_1的输出：$O_{1out} = \frac{1}{1+e^{-O_{1in}}} = 0.7513$

同理，O_2的输出 $O_{2out} = 0.7729$

此时前向传播结束，得到的输出为(0.7513, 0.7729)与(0.01, 0.99)相差甚远

所以需要反向传播，更新权值

损失函数

$$E = \sum \frac{1}{2}(target - O_{out})^2$$

$$= \frac{1}{2}(0.01 - 0.7513)^2 + \frac{1}{2}(0.99 - 0.7729)^2 = 0.2983$$

最终目的就是使E尽量小。

(1). 输出层到隐含层

例如现在我们要更新w_5的权重，需要知道E在w_5上的梯度方向，所以对w_5求偏导

$$\frac{dE}{dw_5} = \frac{dE}{dO_{1out}} \cdot \frac{dO_{1out}}{dO_{1in}} \cdot \frac{dO_{1in}}{dw_5}$$

链式法则，这是反向传播的根本，
同时也表明损失函数和激活函数必须可导

$$\frac{dE}{dO_{1out}} = \frac{d\sum\frac{1}{2}(target - O_{out})^2}{dO_{1out}} = 2 \cdot \frac{1}{2}(target - O_{1out})^{2-1} \cdot (-1) = 0.7413$$

$$\frac{dO_{1out}}{dO_{1in}} = \frac{d\frac{1}{1+e^{-O_{1in}}}}{dO_{1in}} = 0.1868$$ (sigmoid的导数)

$$\frac{dO_{1in}}{dw_5} = \frac{d\, w_5 \cdot h_{1out} + w_6 \cdot h_{2out} + b_2}{dw_5} = h_{1out} = 0.5932$$

最终结果 $\frac{dE}{dw_5} = 0.08216$

若设学习率 $\eta = 0.5$

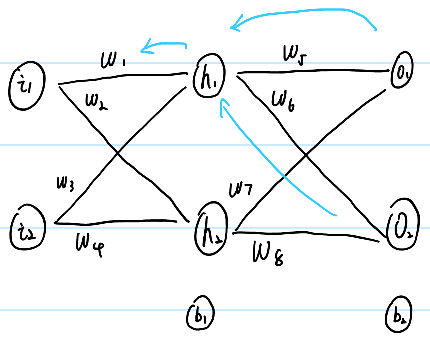
若设学习率 $\eta = 0.5$

则更新后的 $w_5^+ = w_5 - \eta \cdot \frac{\partial E}{\partial w_5} = 0.4 - 0.5 \times 0.08216 = 0.3589$

同理可更新 $w_6^+ = 0.4086$ $w_7^+ = 0.5113$ $w_8^+ = 0.5613$

(2) 隐含层到输入层

原理与之前相同. 区别是. 上面更新 w_5 时. 只要求从 $O_1 \to h_1$ 的路径即可. 但更新一层的 w_1 还受 O_2 影响



w_5 只受 O_1 影响

但 w_1 受后面的所有神经元影响

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial h_{1out}} \cdot \frac{\partial h_{1out}}{\partial h_{1in}} \cdot \frac{\partial h_{1in}}{\partial w_1}$$

$$\frac{\partial E}{\partial h_{1out}} = \frac{\partial \sum \frac{1}{2}(target - O_{out})^2}{\partial h_{1out}} = \frac{\partial E}{\partial O_{1out}} \cdot \frac{\partial O_{1out}}{\partial O_{1in}} \cdot \underbrace{\frac{\partial O_{1in}}{\partial h_{1out}}}_{w_5} + \frac{\partial E}{\partial O_{2out}} \cdot \frac{\partial O_{2out}}{\partial O_{2in}} \cdot \underbrace{\frac{\partial O_{2in}}{\partial h_{1out}}}_{w_6}$$

这时候用 w_5 计算时用的是没更新的 w_5

这样计算完成之后 $\frac{\partial E}{\partial w_1} = 0.0004385$

$w_1^+ = w_1 - \eta \cdot \frac{\partial E}{\partial w_1} = 0.1497$

这样即可完成权值更新