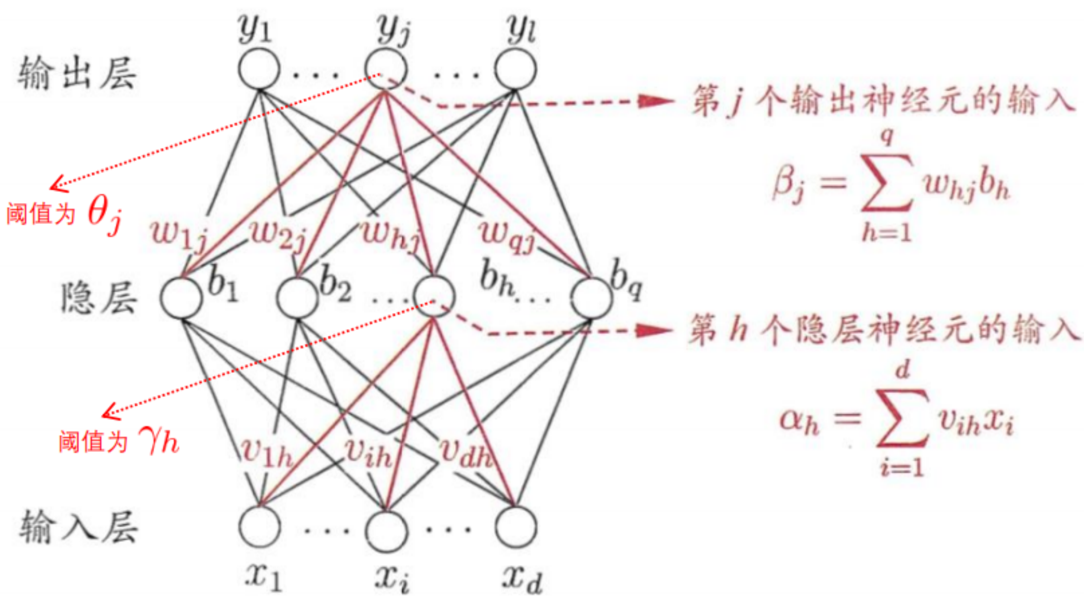
单层前馈神经网络

2022-01-02

1. 单层前馈神经网络



1.1. 符号解释

1. $X^k = \{x_1^k, x_2^k, \cdots, x_d^k\}$：输入的第$k$个数据，共$d$维
2. $V_{d\times q}$：输入层到隐藏层的权重，q为隐藏层结点的个数
3. $\gamma_h, (1 \le h \le q)$：输入层到隐藏层的阈值（偏置）
4. $\alpha_h, (1 \le h \le q)$：隐藏层的输入
5. $b_h, (1 \le h \le q)$：隐藏层的输出
6. $W_{q\times l}$：隐藏层到输出层的权重，l为输出层结点的个数
7. $\theta_j, (1 \le j \le l)$：隐藏层到输出层的阈值（偏置）
8. $\beta_j, (1 \le j \le l)$，输出层的输入
9. $\hat{y}_j^k, (1 \le j \le l)$：输出层的输出
10. $y_j^k, (1 \le j \le l)$：输出层的label
11. f：sigmoid函数，其导数特点为$f'(x) = f(x)(1 - f(x))$
12. E_K：损失函数

1.2. 前馈计算

1. 输入层到隐藏层

$$\begin{aligned} \alpha_h &= \sum_{i=1}^{d} v_{ih} x_i^k \quad \text{其中，} 1 \leq h \leq q \\ b_h &= f(\alpha_h - \gamma_h) \end{aligned} \tag{16}$$

2. 隐藏层到输出层

$$\beta_j = \sum_{h=1}^{q} w_{hj} b_h \hat{y}_j^k = f(\beta_j - \theta_j) \text{其中，} 1 \leq j \leq l \tag{17}$$

3. 损失函数

$$E_k = \frac{1}{2} \sum_{j=1}^{l} \left(\hat{y}_j^k - y_j^k\right)^2 \tag{18}$$

1.3. 反向传播

1. w_{hj}

$$E_k = \frac{1}{2} \sum_{j=1}^{l} \left(\hat{y}_j^k - y_j^k\right)^2$$

$$\downarrow$$

$$\hat{y}_j^k = f(\beta_j - \theta_j) \qquad f \text{ 为Sigmoid函数}$$

$$\downarrow$$

$$\beta_j = \sum_{h=1}^{q} w_{hj} b_h$$

因此，

$$\frac{\partial E_k}{\partial w_{hj}} = \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial w_{hj}} \tag{19}$$

分开计算：

$$\begin{aligned} \frac{\partial E_k}{\partial \hat{y}_j^k} &= \frac{\partial \left[\frac{1}{2} \sum_{j=1}^{l} \left(\hat{y}_j^k - y_j^k\right)^2\right]}{\partial \hat{y}_j^k} \\ &= \frac{1}{2} \times 2 \times \left(\hat{y}_j^k - y_j^k\right) \times 1 \\ &= \hat{y}_j^k - y_j^k \\ \frac{\partial \hat{y}_j^k}{\partial \beta_j} &= \frac{\partial \left[f\left(\beta_j - \theta_j\right)\right]}{\partial \beta_j} \\ &= f'\left(\beta_j - \theta_j\right) \times 1 \\ &= f\left(\beta_j - \theta_j\right) \times \left[1 - f\left(\beta_j - \theta_j\right)\right] \\ &= \hat{y}_j^k \left(1 - \hat{y}_j^k\right) \\ \frac{\partial \beta_j}{\partial w_{hj}} &= \frac{\partial \left(\sum_{h=1}^{q} w_{hj} b_h\right)}{\partial w_{hj}} \\ &= b_h \end{aligned} \tag{20}$$

令:

$$g_j = -\frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} = -\left(\hat{y}_j^k - y_j^k\right) \cdot \hat{y}_j^k \left(1 - \hat{y}_j^k\right) = \hat{y}_j^k \left(1 - \hat{y}_j^k\right)\left(y_j^k - \hat{y}_j^k\right) \tag{21}$$

所以:

$$\begin{aligned} \Delta w_{hj} &= -\eta \frac{\partial E_k}{\partial w_{hj}} \\ &= -\eta \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial w_{hj}} \\ &= \eta g_j b_h \end{aligned} \tag{22}$$

2. θ_j

$$E_k = \frac{1}{2} \sum_{j=1}^{l} \left(\hat{y}_j^k - y_j^k\right)^2$$

$$\downarrow$$

$$\hat{y}_j^k = f(\beta_j - \theta_j)$$

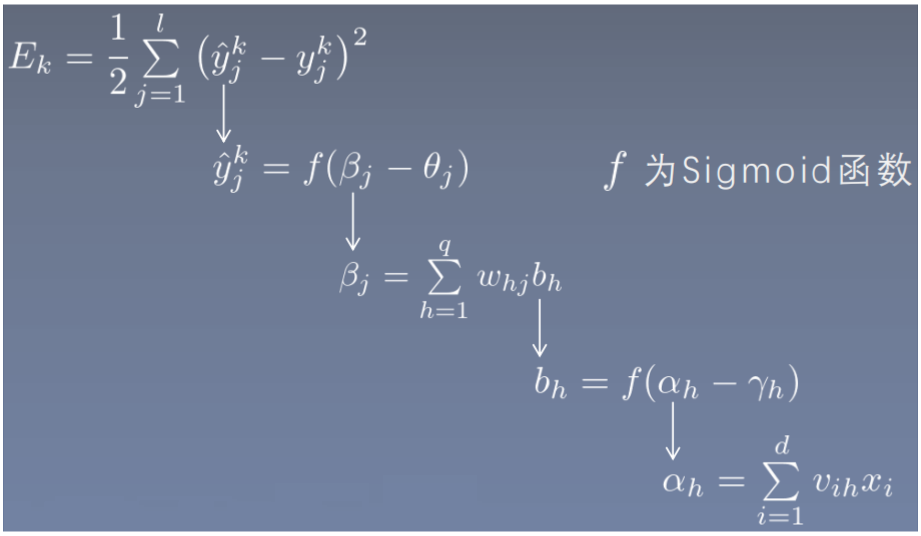
因此,

$$\begin{aligned} \frac{\partial E_k}{\partial \theta_j} &= \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \theta_j} \\ &= \left(\hat{y}_j^k - y_j^k\right) \cdot \frac{\partial \hat{y}_j^k}{\partial \theta_j} \\ &= \left(\hat{y}_j^k - y_j^k\right) \cdot \frac{\partial \left[f\left(\beta_j - \theta_j\right)\right]}{\partial \theta_j} \\ &= \left(\hat{y}_j^k - y_j^k\right) \cdot f'\left(\beta_j - \theta_j\right) \times -1 \\ &= \left(y_j^k - \hat{y}_j^k\right) \cdot f'\left(\beta_j - \theta_j\right) \\ &= \left(y_j^k - \hat{y}_j^k\right) \hat{y}_j^k \left(1 - \hat{y}_j^k\right) \end{aligned} \tag{23}$$

所以:

$$\begin{aligned} \Delta \theta_j &= -\eta \frac{\partial E_k}{\partial \theta_j} \\ &= -\eta \left(y_j^k - \hat{y}_j^k\right) \hat{y}_j^k \left(1 - \hat{y}_j^k\right) \\ &= -\eta g_j \end{aligned} \tag{24}$$

3. v_{ih}



因此，

$$\frac{\partial E_k}{\partial v_{ih}} = \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot \frac{\partial b_h}{\partial \alpha_h} \cdot \frac{\partial \alpha_h}{\partial v_{ih}} \tag{25}$$

分开计算，

$$\begin{aligned}
\frac{\partial \beta_j}{\partial b_h} &= \frac{\partial\left(\sum_{h=1}^{q} w_{hj} b_h\right)}{\partial b_h} \\
&= w_{hj} \\
\frac{\partial b_h}{\partial \alpha_h} &= \frac{\partial\left[f\left(\alpha_h - \gamma_h\right)\right]}{\partial \alpha_h} \\
&= f'\left(\alpha_h - \gamma_h\right) \times 1 \\
&= f\left(\alpha_h - \gamma_h\right) \times \left[1 - f\left(\alpha_h - \gamma_h\right)\right] \\
&= b_h\left(1 - b_h\right) \\
\frac{\partial \alpha_h}{\partial v_{ih}} &= \frac{\partial\left(\sum_{i=1}^{d} v_{ih} x_i\right)}{\partial v_{ih}} \\
&= x_i
\end{aligned} \tag{26}$$

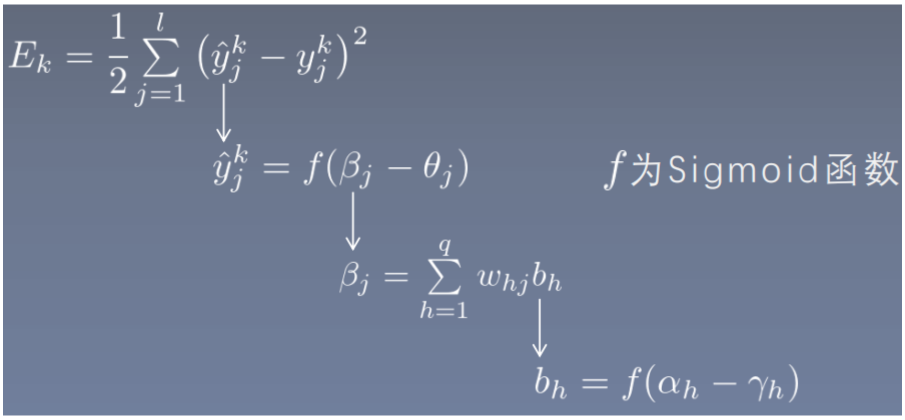
令：

$$e_h = -\frac{\partial E_k}{\partial \alpha_h} = -\sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot \frac{\partial b_h}{\partial \alpha_h} = b_h\left(1 - b_h\right) \sum_{j=1}^{l} w_{hj} g_j \tag{27}$$

所以，

$$\begin{aligned}
\Delta v_{ih} &= -\eta \frac{\partial E_k}{\partial v_{ih}} \\
&= -\eta \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot \frac{\partial b_h}{\partial \alpha_h} \cdot \frac{\partial \alpha_h}{\partial v_{ih}} \\
&= \eta e_h x_i
\end{aligned} \tag{28}$$

4. γ_h



因此,

$$
\begin{aligned}
\frac{\partial E_k}{\partial \gamma_h} &= \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot \frac{\partial b_h}{\partial \gamma_h} \\
&= \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot \frac{\partial \left[f\left(\alpha_h - \gamma_h\right)\right]}{\partial \gamma_h} \\
&= \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot f'\left(\alpha_h - \gamma_h\right) \cdot (-1) \\
&= \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot f\left(\alpha_h - \gamma_h\right) \times \left[1 - f\left(\alpha_h - \gamma_h\right)\right] \cdot (-1) \\
&= \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot b_h\left(1 - b_h\right) \cdot (-1) \\
&= \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot w_{hj} \cdot b_h\left(1 - b_h\right) \cdot (-1) \\
&= \sum_{j=1}^{l} g_j \cdot w_{hj} \cdot b_h\left(1 - b_h\right) \\
&= e_h
\end{aligned} \tag{29}
$$

所以:

$$
\begin{aligned}
\Delta \gamma_h &= -\eta \frac{\partial E_k}{\partial \gamma_h} \\
&= -\eta e_h
\end{aligned} \tag{30}
$$