误差反传算法

邢超 (xingnix@live.com)

September 26, 2016

1 向量描述

1.1 $\Delta E, \Delta O, \Delta W$

对于第 i 层网络：

$$\begin{aligned}\vec{O}_i &= F_i(W_i\vec{X}_i + \vec{b}_i) \\ &= \begin{bmatrix} f_i(\vec{w}_1\vec{X}_i + b_{i,1}) & f_i(\vec{w}_2\vec{X}_i + b_{i,2}) & \cdots & f_i(\vec{w}_n\vec{X}_i + b_{i,n}) \end{bmatrix}^T\end{aligned}$$

其中：

- $\vec{O}_i$ 为第 i 层网络输出，
- $\vec{X}_i$ 为第 i 层网络输入，等于第 $i-1$ 层网络输出 $\vec{O}_{i-1}$
- $\vec{w}_j$ 为第 i 层网络权值矩阵 W_i 的第 j 行，
- $b_{i,j}$ 为向量 $\vec{b}_i$ 的第 j 个元素。

对于 W_i 的一个微小变化 ΔW_i ，得：

$$\begin{aligned}\Delta\vec{O}_i &= \begin{bmatrix} f_i'(\vec{w}_1\vec{X}_i + b_{i,1})\Delta\vec{w}_1\vec{X}_i & f_i'(\vec{w}_2\vec{X}_i + b_{i,2})\Delta\vec{w}_2\vec{X}_i & \cdots & f_i'(\vec{w}_n\vec{X}_i + b_{i,n})\Delta\vec{w}_n\vec{X}_i \end{bmatrix}^T \\ &= diag[F_i'(W_i\vec{X}_i + \vec{b}_i)]\Delta W\vec{X}_i \\ &= diag(F_i')\Delta W_i\vec{X}_i\end{aligned}$$

其中 $diag(F)$ 表示主对角元素为向量 F 的元素的方阵。
对于 N 层网络的误差函数：

$$E = (\vec{Y} - \vec{O}_N)^T(\vec{Y} - \vec{O}_N)$$

误差函数的增量：

$$\begin{aligned}\Delta E &= -2(\vec{Y} - \vec{O}_N)^T \cdot \Delta\vec{O}_N \\ &= -2(\vec{Y} - \vec{O}_N)^T \cdot diag(F_N')\Delta W_N\vec{X}_N \\ \frac{\partial E}{\partial W_N} &= -2diag(F_N') \cdot (\vec{Y} - \vec{O}_N) \cdot \vec{X}_N^T\end{aligned}$$

1.2 反向传播

根据链式法则：

$$\begin{aligned}\Delta\vec{O}_i &= \begin{bmatrix} f_i'(\vec{w}_1\vec{X}_i + b_{i,1})\vec{w}_1\Delta\vec{X}_i & f_i'(\vec{w}_2\vec{X}_i + b_{i,2})\vec{w}_2\Delta\vec{X}_i & \cdots & f_i'(\vec{w}_n\vec{X}_i + b_{i,n})\vec{w}_n\Delta\vec{X}_i \end{bmatrix}^T \\ &= diag(F_i')W_i\Delta\vec{X}_i \\ \Delta E &= 2(\vec{O}_N - \vec{Y})^T \cdot diag(F_N') \cdot W_N \cdot \Delta\vec{X}_N \\ &= 2(\vec{O}_N - \vec{Y})^T \cdot diag(F_N') \cdot W_N \cdot \Delta\vec{O}_{N-1} \\ &= 2(\vec{O}_N - \vec{Y})^T \cdot diag(F_N') \cdot W_N \cdot diag(F_{N-1}') \cdot \Delta W_N \cdot \vec{X}_{N-1}\end{aligned}$$

对于 N 层网络可得：

$$\begin{aligned}\Delta E &= \vec{\alpha}^T \cdot \Delta W_i \cdot \vec{\beta} \\ \frac{\partial E}{\partial W_i} &= \vec{\alpha} \cdot \vec{\beta}^T\end{aligned}$$

其中：

$$\begin{aligned}\vec{\alpha}^T &= 2(\vec{O}_N - \vec{Y})^T \cdot diag(F_N') \cdot \left[\prod_{n=i+1}^{N} W_n \cdot diag(F_{n-1}')\right] \\ \vec{\beta} &= \vec{X}_i \qquad \text{(inputs of the i'th layer)}\end{aligned}$$

反向传播：

$$\begin{aligned}\vec{\delta}_N &= [2(\vec{Y}_o - \vec{Y}_E)^T \cdot diag(F_N')]^T \\ &= diag(F_N') \cdot 2(\vec{Y}_o - \vec{Y}_E) \\ \vec{\delta}_n &= diag(F_n') \cdot W_{n+1}^T \cdot \vec{\delta}_{n+1} \\ \Delta W_i &= \delta_i \cdot \vec{X}_i^T\end{aligned}$$

1.3 偏置向量

对偏置向量的更新可考虑增广权植矩阵 $[W|b]$，增广输入 $\begin{bmatrix}\vec{X}_o \\ 1\end{bmatrix}$ 。令 $\beta = \begin{bmatrix}\vec{X}_o \\ 1\end{bmatrix}$ 即可得到增广权值矩阵的更新公式。

2 递推描述

2.1 符号表示

考虑神经元 q ，有：

$$
\begin{aligned}
q_x &= \sum_{p\in\{p'|p'\to q\}} w_{qp} p_o \\
q_o &= f(q_x)
\end{aligned}
$$

其中：

- q_x 表示神经元 q 的输入
- q_o 表示神经元 q 的输出
- p_o 表示神经元 p 的输出
- $f(\cdot)$ 表示神经元的非线性函数
- w_{qp} 表示连接神经元 q 输出端到 p 输入端的网络权值
- $\{p'|p'\to q\}$ 表示神经元 p' 的集合，p' 的输出端连接到 q 输入端的

2.2 输出层神经元 t：

$$
\begin{aligned}
\frac{\partial E}{\partial t_x} &= \frac{\partial E}{\partial t_o}\frac{\partial t_o}{\partial t_x} \\
&= \frac{\partial}{\partial t_o}[\frac{1}{2}(t_o - t_y)^2]\frac{\partial t_o}{\partial t_x} \\
&= (t_o - t_y)\frac{\partial t_o}{\partial t_x} \\
&= (t_o - t_y)f'(t_x) \\
\frac{\partial E}{\partial w_{tq}} &= \frac{\partial E}{\partial t_x}\frac{\partial t_x}{\partial w_{tq}} \\
&= \frac{\partial E}{\partial t_x} q_o
\end{aligned}
$$

其中：

- t_y 为输出神经元 t 的期望输出
- $f'(t_x) = \frac{df(t_x)}{dt_x}\Big|_{t_x}$
- q 为输出端连接到 t 的输入端的某个神经元

2.3 其它神经元 q:

$$
\begin{aligned}
\frac{\partial E}{\partial q_x} &= \sum_{r \in \{r' | q \to r'\}} \frac{\partial E}{\partial r_x} \frac{\partial r_x}{\partial q_x} \\
&= \sum_{r \in \{r' | q \to r'\}} \frac{\partial E}{\partial r_x} \frac{\partial r_x}{\partial q_o} \frac{\partial q_o}{\partial q_x} \\
&= \sum_{r \in \{r' | q \to r'\}} \frac{\partial E}{\partial r_x} w_{rq} f'(q_x) \\
\frac{\partial E}{\partial w_{qp}} &= \frac{\partial E}{\partial q_x} \frac{\partial q_x}{\partial w_{qp}} \\
&= \frac{\partial E}{\partial q_x} p_o
\end{aligned}
$$

2.4 反向传播

输出层神经元：

$$
\begin{aligned}
t_\delta &= \frac{\partial E}{\partial t_x} \\
&= (t_o - t_y) f'(t_x) \\
\Delta w_{tq} &= -\eta t_\delta q_o
\end{aligned}
$$

其它神经元：

$$
\begin{aligned}
q_\delta &= \frac{\partial E}{\partial q_x} \\
&= \sum_{\{r | q \to r\}} r_\delta w_{rq} f'(q_x) \\
\Delta w_{qp} &= -\eta q_\delta p_o
\end{aligned}
$$