误差反传算法

邢超 (xingnix@live.com)

September 19, 2016

1 ΔW 计算

误差函数：

$$E = (\vec{Y_E} - \vec{Y_o})^T (\vec{Y_E} - \vec{Y_o})$$

每层输出：

$$\begin{aligned}
\vec{X_o} &= F(W\vec{X_i} + \vec{b}) \\
&= \begin{bmatrix} f(\vec{w}_1\vec{X_i} + b_1) & f(\vec{w}_2 x_i + b_2) & \cdots & f(\vec{w}_n\vec{X_i} + b_n) \end{bmatrix}^T
\end{aligned}$$

其中 $\vec{w_i}$ 为 W 矩阵的第 i 行。

对于 W 的一个微小变化 ΔW ，得：

$$\begin{aligned}
\Delta\vec{X_o} &= \begin{bmatrix} f'(\vec{w}_1\vec{X_i} + b_1)\Delta\vec{w}_1\vec{X_i} & f'(\vec{w}_2\vec{X_i} + b_2)\Delta\vec{w}_2\vec{X_i} & \cdots & f'(\vec{w}_n\vec{X_i} + b_n)\Delta\vec{w}_n\vec{X_i} \end{bmatrix}^T \\
&= diag[F'(W\vec{X_i} + \vec{b})]\Delta W\vec{X_i} \\
&= diag(F')\Delta W\vec{X_i}
\end{aligned}$$

其中 $diag(F)$ 表示主对角元素为向量 F 的元素的方阵。

误差函数的增量：

$$\begin{aligned}
\Delta E = \Delta E &= -2(\vec{Y_E} - \vec{Y_o})^T \cdot diag(F')\Delta(W\vec{X_o} + b) \\
&= -2(\vec{Y_E} - \vec{Y_o})^T \cdot diag(F')\Delta W\vec{X_o}
\end{aligned}$$

2 反向传播

根据链式法则：

$$\begin{aligned}
\Delta\vec{X_o} &= diag(F')\Delta(W\vec{X_i} + \vec{b}) \\
&= diag(F')W\Delta\vec{X_i} \\
\Delta E &= 2(\vec{Y_o} - \vec{Y_E})^T \cdot diag(F') \cdot W \cdot \Delta\vec{X_o} \\
&= 2(\vec{Y_o} - \vec{Y_E})^T \cdot diag(F') \cdot W \cdot diag(F') \cdot \Delta W \cdot \vec{X_i}
\end{aligned}$$

对于多层网络权值的更新：

$$\Delta E = \vec{a}^T \cdot \Delta W \cdot \vec{b}$$
$$\frac{\partial E}{\partial W} = \vec{a} \cdot \vec{b}^T$$

其中：

$$\vec{a}^T = 2(\vec{Y}_o - \vec{Y}_E)^T \cdot \prod_{n=1}^{N} [diag(F') \cdot W] \cdot diag(F') \qquad (n = 0, 1, 2, \ldots)$$
$$\vec{b} = \vec{X}_i \qquad (\text{ inputs of the layer})$$

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 $b = \begin{bmatrix} \vec{X}_o \\ 1 \end{bmatrix}$ 即可得到增广权值矩阵的更新公式。