

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:

$$\begin{aligned}\vec{X}_o &= F(W\vec{X}_i + \vec{b}) \\ &= [f(\vec{w}_1\vec{X}_i + b_1) \quad f(\vec{w}_2\vec{X}_i + b_2) \quad \cdots \quad f(\vec{w}_n\vec{X}_i + b_n)]^T\end{aligned}$$

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:

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

其中 $\text{diag}(F)$ 表示主对角元素为向量 F 的元素的方阵。

误差函数的增量:

$$\begin{aligned}\Delta E &= \Delta E = -2(\vec{Y}_E - \vec{Y}_o)^T \cdot \text{diag}(F')\Delta(W\vec{X}_o + \vec{b}) \\ &= -2(\vec{Y}_E - \vec{Y}_o)^T \cdot \text{diag}(F')\Delta W\vec{X}_o\end{aligned}$$

2 反向传播

根据链式法则:

$$\begin{aligned}\Delta\vec{X}_o &= \text{diag}(F')\Delta(W\vec{X}_i + \vec{b}) \\ &= \text{diag}(F')W\Delta\vec{X}_i \\ \Delta E &= 2(\vec{Y}_o - \vec{Y}_E)^T \cdot \text{diag}(F') \cdot W \cdot \Delta\vec{X}_o \\ &= 2(\vec{Y}_o - \vec{Y}_E)^T \cdot \text{diag}(F') \cdot W \cdot \text{diag}(F') \cdot \Delta W \cdot \vec{X}_i\end{aligned}$$

多层网络:

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

其中:

$$\begin{aligned}\vec{\alpha}^T &= 2(\vec{Y}_o - \vec{Y}_E)^T \cdot \prod_{n=i+1}^N [\text{diag}(F') \cdot W] \cdot \text{diag}(F') \\ \vec{\beta} &= \vec{X}_i \quad (\text{inputs of the layer})\end{aligned}$$

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