MLP_BP 权重和偏置推导

2010239 李思凡

1. 对损失函数求偏导：

已知损失函数为：

$$\mathrm{E}=\sum_{\mathrm{i}}\sum_{\mathrm{j}}\frac{1}{2}(y_{ij}-d_{ij})^2+\frac{1}{2}\gamma(\mathrm{tr}(S_w)-\mathrm{tr}(S_b))$$

需要求出损失函数对y_{ij}的偏导，因此需要首先表示出 $\mathrm{tr}(S_w)$和 $\mathrm{tr}(S_b)$。

已知类内散度和类间散度S_w, S_b的定义为：

$$S_w=\sum_{c=1}^{C}\sum_{{y_i}^M\in c}({y_i}^M-{m_c}^M)({y_i}^M-{m_c}^M)^T$$

$$S_b=\sum_{c=1}^{C}n_c({m_c}^M-m^M)({m_c}^M-m^M)^T$$

由于条件中给出的损失函数中使用了上述两个矩阵的迹，因此将${m_c}^M$和m^M表示出来，并计算矩阵的迹为：

$$\mathrm{tr}(S_w)=\sum_{c=1}^{C}\sum_{{y_i}^M\in c}\sum_{j=1}^{M}(y_{ij}-\frac{1}{n_c}\sum_{{y_k}^M\in c}y_{k,j})^2$$

$$\mathrm{tr}(S_b)=\sum_{c=1}^{C}n_c\sum_{j=1}^{M}(\frac{1}{n_c}\sum_{{y_k}^M\in c}y_{k,j}-\frac{\sum_{c=1}^{C}\sum_{{y_k}^M\in c}y_{k,j}}{\sum_{c=1}^{C}n_c})^2$$

计算二者对于y_{ij}的偏导为：

$$\frac{\partial\mathrm{tr}(S_w)}{\partial y_{ij}}=2(y_{ij}-\frac{1}{n_c}\sum_{{y_k}^M\in c}y_{k,j})-\frac{2}{n_c}\sum_{{y_i}^M\in c}(y_{ij}-\frac{1}{n_c}\sum_{{y_k}^M\in c}y_{k,j})=2(y_{ij}-m_{cj})$$

$$\begin{aligned}\frac{\partial\mathrm{tr}(S_b)}{\partial y_{ij}}&=2(\frac{1}{n_c}\sum_{{y_k}^M\in c}y_{k,j}-\frac{\sum_{c=1}^{C}\sum_{{y_k}^M\in c}y_{k,j}}{\sum_{c=1}^{C}n_c})\\&\quad+2\sum_{c=1}^{C}n_c(\frac{1}{n_c}\sum_{{y_k}^M\in c}y_{k,j}-\frac{\sum_{c=1}^{C}\sum_{{y_k}^M\in c}y_{k,j}}{\sum_{c=1}^{C}n_c})(-\frac{1}{\sum_{c=1}^{C}n_c})\\&=2(m_{cj}-m_j)\end{aligned}$$

因此，损失函数对y_{ij}的偏导为：

$$\frac{\partial E}{\partial y_{ij}} = (y_{ij} - d_{ij}) + \gamma(y_{ij} + m_j - 2m_{cj})$$

用矩阵可以表示为：

$$\boldsymbol{\frac{\partial E}{\partial y^M} = (y^M - d) + \gamma(y^M + m^M - 2{m_c}^M)}$$

2. 输出时：

输出y^M是其输入的线性组合，假设${x_3}^N$是其输入，此时权重就为 M × N 的矩阵${W_3}^{\mathrm{M\times N}}$，偏置为 N × 1 的矩阵${b_3}^{\mathrm{N}}$。输出为$y^M$=${W_3}^{\mathrm{M\times N}}{x_3}^N + {b_3}^{\mathrm{N\times 1}}$。

继续执行链式求导：

$$\frac{\partial E}{\partial {W_3}^{M\times N}} = \sum_{i=1}^{N} \frac{\partial E}{\partial y_i}\frac{y_i}{\partial {W_3}^{M\times N}} = \frac{\partial E}{\partial y^M}({x_3}^N)^T$$

$$\frac{\partial E}{\partial {b_3}^M} = \frac{\partial E}{\partial y^M}$$

$$\frac{\partial E}{\partial {x_3}^N} = \sum_{i=1}^{N} \frac{\partial E}{\partial y_i}\frac{y_i}{\partial {x_3}^N} = \frac{\partial E}{\partial y^M}({W_3}^{M\times N})^T$$

3. 隐藏层：

隐藏层是在输入的线性组合的基础上，添加了激活函数。采用 sigmoid 激活函数 $\mathrm{y} = \frac{1}{1+\mathrm{e}^{-\mathrm{x}}}$ 实现非线性。输入为${x_l}^N$，权重为${W_l}^{\mathrm{M\times N}}$，偏置为${b_l}^{\mathrm{N}}$。隐藏层的输出为${z_l}^M = sigmoid({y_l}^M) = sigmoid({W_l}^{\mathrm{M\times N}}{x_l}^N + {b_l}^{\mathrm{N\times 1}})$。

继续执行链式求导：

$$\frac{\partial z^M}{\partial y^M} = z^M \otimes (1 - z^M)$$

$$\frac{\partial E}{\partial {W_l}^{M\times N}} = \frac{\partial E}{\partial {z_l}^M} \otimes {z_l}^M \otimes (1 - {z_l}^M)({x_l}^N)^T$$

$$\frac{\partial E}{\partial {b_l}^N} = \frac{\partial E}{\partial {z_l}^M} \otimes {z_l}^M \otimes (1 - {z_l}^M)$$

由于当前层的输出是下一层的输入，有$\frac{\partial E}{\partial {z_l}^M} = \frac{\partial E}{\partial {x_{l+1}}^{Ml+1}}$。

由于是三层感知机，从输入层到隐藏层和从隐藏层到输出层的计算均为上述

过程。

4. 整合 BP 过程

假设输入层输入为$x_1{}^{N_1}$，隐藏层输入为$x_2{}^{N_2}$，输出层输入为$x_3{}^{N_3}$，输出层输出为y^M。使用链式求导整合整个过程。

输出层$W_3{}^{\mathrm{M}\times N_3}$和$b_3{}^M$的梯度为：

$$\frac{\partial E}{\partial W_3{}^{\mathrm{M}\times N_3}}=\frac{\partial E}{\partial y^M}(x_3{}^{N_3})^T=[(y^M-d)+\gamma(y^M+m^M-2m_c{}^M)](x_3{}^{N_3})^T$$

$$\frac{\partial E}{\partial b_3{}^M}=\frac{\partial E}{\partial y^M}=(y^M-d)+\gamma(y^M+m^M-2m_c{}^M)$$

隐藏层$W_2{}^{N_3\times N_2}$和$b_2{}^{N_3}$的梯度为：

$$\begin{aligned}\frac{\partial E}{\partial W_2{}^{N_3\times N_2}}&=\frac{\partial E}{\partial z_2{}^{N_3}}\otimes z_2{}^{N_3}\otimes(1-z_2{}^{N_3})(x_2{}^{N_2})^T=\frac{\partial E}{\partial x_3{}^{N_3}}\otimes x_3{}^{N_3}\otimes(1-x_3{}^{N_3})(x_2{}^{N_2})^T\\&=\frac{\partial E}{\partial y^M}(W_3{}^{M\times N_3})^T\otimes x_3{}^{N_3}\otimes(1-x_3{}^{N_3})(x_2{}^{N_2})^T\\&=[(y^M-d)+\gamma(y^M+m^M-2m_c{}^M)](W_3{}^{M\times N_3})^T\otimes x_3{}^{N_3}\otimes(1-x_3{}^{N_3})(x_2{}^{N_2})^T\end{aligned}$$

$$\begin{aligned}\frac{\partial E}{\partial b_2{}^{N_3}}&=\frac{\partial E}{\partial z_2{}^{N_3}}\otimes z_2{}^{N_3}\otimes(1-z_2{}^{N_3})=\frac{\partial E}{\partial x_3{}^{N_3}}\otimes x_3{}^{N_3}\otimes(1-x_3{}^{N_3})\\&=[(y^M-d)+\gamma(y^M+m^M-2m_c{}^M)](W_3{}^{M\times N_3})^T\otimes x_3{}^{N_3}\otimes(1-x_3{}^{N_3})\end{aligned}$$

输入层$W_1{}^{N_2\times N_1}$和$b_1{}^{N_2}$的梯度为：

$$\begin{aligned}\frac{\partial E}{\partial W_1{}^{N_2\times N_1}}&=\frac{\partial E}{\partial z_1{}^{N_2}}\otimes z_1{}^{N_2}\otimes(1-z_1{}^{N_2})(x_1{}^{N_1})^T=\frac{\partial E}{\partial x_2{}^{N_2}}\otimes x_2{}^{N_2}\otimes(1-x_2{}^{N_2})(x_1{}^{N_1})^T\\&=(W_2{}^{N_3\times N_2})^T[\frac{\partial E}{\partial y^M}(W_3{}^{M\times N_3})^T\otimes x_3{}^{N_3}\otimes(1-x_3{}^{N_3})]\otimes x_2{}^{N_2}\otimes(1-x_2{}^{N_2})(x_1{}^{N_1})^T\\&=(W_2{}^{N_3\times N_2})^T\{[(y^M-d)+\gamma(y^M+m^M-2m_c{}^M)](W_3{}^{M\times N_3})^T\otimes x_3{}^{N_3}\otimes(1\\&\qquad-x_3{}^{N_3})\}\otimes x_2{}^{N_2}\otimes(1-x_2{}^{N_2})(x_1{}^{N_1})^T\end{aligned}$$

$$\frac{\partial E}{\partial b_1{}^{N_2}}=\frac{\partial E}{\partial z_1{}^{N_2}}\otimes z_1{}^{N_2}\otimes(1-z_1{}^{N_2})=\frac{\partial E}{\partial x_2{}^{N_2}}\otimes x_2{}^{N_2}\otimes(1-x_2{}^{N_2})$$

$$\begin{aligned}&=(W_2{}^{N_3\times N_2})^T\{[(y^M-d)+\gamma(y^M+m^M-2m_c{}^M)](W_3{}^{M\times N_3})^T\otimes x_3{}^{N_3}\otimes(1\\&\qquad-x_3{}^{N_3})\}\otimes x_2{}^{N_2}\otimes(1-x_2{}^{N_2})\end{aligned}$$

因此，更新时相应权重和偏置为：

$$W_l^{'} = W_l - \eta \frac{\partial E}{\partial W_l}$$
$$b_l^{'} = b_l - \eta \frac{\partial E}{\partial b_l}$$