MLP_BP实验报告

【网络结构】

- 输入层：

  输入向量：$\boldsymbol{x}$，维度为S
- 系数矩阵W'：$S \times M$
- 隐藏层：

  激活函数：`tanh`函数
- 系数矩阵W'：$M \times N$
- 输出层：

  输出向量：$\boldsymbol{y}$，维度为M

一、损失函数求导

【约定】

- n_c表示类别为c的样本数
- N表示样本总数

损失函数：

$$E=\sum_i\sum_j\frac{1}{2}(y_{i,j}^M-d_{i,j})^2+\frac{1}{2}\gamma(tr(S_w)-tr(S_b))$$

分部求导：

- 对第一项求导：

$$\frac{\partial[\sum_i\sum_j\frac{1}{2}(y_{i,j}^M-d_{i,j})^2]}{\partial y_{i,j}}=y_{i,j}-d_{i,j}$$

- 对$tr(S_w)$求导：

  展开$tr(S_w)$：

$$tr(S_w)=\sum_{c=1}^C\sum_{y_i\in c}\sum_j(y_{i,j}-m_{c,j})^2$$

求导：

$$
\begin{aligned}
\frac{\partial\, \mathrm{tr}(S_w)}{\partial y_{ij}} &= \left(2-\frac{1}{n_c}\right)(y_{ij}-m_{ij})+2\cdot\left(-\frac{1}{n_c}\right)\sum_{k\neq i}^{K}(y_{kj}-m_{cj}) \\
&= 2\,(y_{ij}-m_{ij})+2\cdot\left(-\frac{1}{n_c}\right)\sum_{y_{i\in C}}(y_{ij}-m_{cj}) \\
&= 2\,(y_{ij}-m_{cj})
\end{aligned}
$$

- 对$tr(S_b)$求导：

  展开$tr(S_b)$：

$$
tr(S_b)=\sum_{c=1}^{C}\sum_{j}(m_{cj}-m_j)^2
$$

  求导：

$$
\begin{aligned}
\frac{\partial tr\,(S_b)}{\partial y_{ij}} &= 2n_c\,(m_{cj}-m_j)\left(\frac{1}{n_c}-\frac{1}{N}\right)+2n_c\cdot\left(-\frac{1}{N}\right)\sum_{R\neq j}^{k}(m_{kj}-m_j) \\
&= 2\,(m_{cj}-m_j)+2n_c\cdot\left(-\frac{1}{N}\right)\sum_{y_i\in C}(m_{ij}-m_j) \\
&= 2\,(m_{cj}-m_j)
\end{aligned}
$$

- 综上，

  损失函数对第i个预测向量的某一项求导：

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

损失函数对第i个预测值求导：

$$
\frac{\partial E}{\partial \boldsymbol{y_i}}=(\boldsymbol{y_i}-\boldsymbol{d_i})+\gamma(\boldsymbol{y_i}+\boldsymbol{m}-2\boldsymbol{m_c})
$$

二、输出层求导

1. 输出层表达式：

$$
\boldsymbol{y}^M=W^{M\times N}\cdot\boldsymbol{z}^N+\boldsymbol{b}^M
$$

2. 损失函数E对系数矩阵W求导：

   对某一项求导：

$$
\frac{\partial E}{\partial w_{ij}}=\frac{\partial E}{\partial y_j}\cdot\frac{\partial y_i}{\partial w_{ij}}=z_j
$$

可知E对W求导：

$$
\frac{\partial E}{\partial W} = \begin{bmatrix} \frac{\partial E}{y_1} \cdot \frac{\partial y_1}{\partial w_{11}} & \cdots & \frac{\partial E}{\partial y_1} \cdot \frac{\partial y_1}{\partial w_{mn}} \\ \frac{\partial E}{y_2} \cdot \frac{\partial y_2}{\partial u_{21}} & \ddots & \vdots \\ \vdots & & \\ \frac{\partial E}{y_m} \cdot \frac{\partial y_m}{\partial w_{m1}} & \frac{\partial E}{y_m} \cdot \frac{\partial y_m}{\partial w_{m2}} & \frac{\partial E}{y_m} \cdot \frac{\partial y_m}{\partial w_{mn}} \end{bmatrix} = \frac{\partial E}{\partial \boldsymbol{y}} \cdot \boldsymbol{z}^T
$$

3. 损失函数E对偏置项b求导：

   对某一项求导：

$$
\frac{\partial E}{\partial b_i} = \frac{\partial E}{\partial y_i} \cdot \frac{\partial y_i}{\partial b_i} = \frac{\partial E}{\partial y_i}
$$

可知E对b求导：

$$
\frac{\partial E}{\partial \boldsymbol{b}} = \frac{\partial E}{\partial \boldsymbol{y}} \cdot \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{b}} = \frac{\partial E}{\partial \boldsymbol{y}}
$$

4. 损失函数E对输入向量$\boldsymbol{z}$求导：

   对某一项求导：

$$
\frac{\partial E}{\partial z_j} = \sum_{i}^{M} \frac{\partial E}{\partial y_i} \cdot \frac{\partial y_i}{\partial z_j} = \sum_{i}^{M} \frac{\partial E}{\partial y_i} w_{ij}
$$

可知E对$\boldsymbol{z}$求导：

$$
\frac{\partial E}{\partial \boldsymbol{z}} = W^T \cdot \frac{\partial E}{\partial \boldsymbol{y}}
$$

三、隐藏层求导

本题未说明激活函数，但其对网络的非线性十分重要，所以我选择使用`tanh`函数作为隐藏层的激活函数

设隐藏层输入向量为$\boldsymbol{t}$

即：

$$
\boldsymbol{z} = tanh(\boldsymbol{t})
$$

隐藏层求导：

$$
\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{t}} = 1 - \boldsymbol{z}^2
$$

注：其中$\boldsymbol{z}^2$表示$\boldsymbol{z}$对位平方运算得到的向量。

四、输入层求导

设输入向量为$\boldsymbol{x}$，维度为S，则表达式、系数矩阵求导、偏置项求导均类似输出层。

表达式：

$$\boldsymbol{t}^N = W'^{N\times S}\cdot\boldsymbol{x}^S + \boldsymbol{b'}^N$$

系数矩阵W'求导：

$$\frac{\partial E}{\partial W'} = \frac{\partial E}{\partial \boldsymbol{z}}\odot\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{t}}\cdot\boldsymbol{x}^T$$

注：其中⊙表示向量对位相乘

偏置项$\boldsymbol{b}'$求导：

$$\frac{\partial E}{\partial b'} = \frac{\partial E}{\partial \boldsymbol{z}}\odot\frac{\partial z}{\partial \boldsymbol{t}}$$

五、最终结果计算

将中间计算结果逐层带入，得到最终结果：

- 隐藏层 -> 输出层：

$$\frac{\partial E}{\partial W} = [(\boldsymbol{y_i}-\boldsymbol{d_i})+\gamma(\boldsymbol{y_i}+\boldsymbol{m}-2\boldsymbol{m_c})]\cdot\boldsymbol{z}^T$$

$$\frac{\partial E}{\partial \boldsymbol{b}} = (\boldsymbol{y_i}-\boldsymbol{d_i})+\gamma(\boldsymbol{y_i}+\boldsymbol{m}-2\boldsymbol{m_c})$$

- 输入层 -> 隐藏层：

$$\frac{\partial E}{\partial W'} = W^T\cdot[(\boldsymbol{y_i}-\boldsymbol{d_i})+\gamma(\boldsymbol{y_i}+\boldsymbol{m}-2\boldsymbol{m_c})]\odot(1-\boldsymbol{z}^2)\cdot\boldsymbol{x}^T$$

$$\frac{\partial E}{\partial b'} = W^T\cdot[(\boldsymbol{y_i}-\boldsymbol{d_i})+\gamma(\boldsymbol{y_i}+\boldsymbol{m}-2\boldsymbol{m_c})]\odot(1-\boldsymbol{z}^2)$$

求出导数后，每次进行梯度下降即可。

六、总结

通过本次实验，我对多层感知机有了更深刻的理解，

通过手算梯度对反向传递算法有了更深刻的认识。