计算机学院

机器学习导论实验报告

作业二
三层 MLP 反向传播

姓名：赵健坤

学号：2010535

专业：计算机科学与技术

2022 年 11 月 11 日

目录

1 损失函数求导

首先，将 $\mathrm{tr}\left(S_w\right)$ 和 $\mathrm{tr}\left(S_b\right)$ 表示为 $y_{i,j}^M$ 的函数：

$$
\begin{aligned}
\mathrm{tr}\left(S_w\right) &= \sum_{c=1}^{C} \sum_{y_i^M \in c} \sum_{j} \left(y_{i,j}^M - m_{c,j}^M\right)^2 \\
&= \sum_{c=1}^{c} \sum_{y_i^M \in c} \sum_{j} \left(y_{i,j}^M - \frac{\sum_{y_k^m \in c^M} y_{k,j}}{n_c}\right)^2 \\
\mathrm{tr}\left(S_b\right) &= \sum_{c=1}^{C} n_c \sum_{j} \left(m_{c,j}^M - m_j^m\right)^2 \\
&= \sum_{c=1}^{c} n_c \sum_{j} \left(\frac{\sum_{y_k^M \in c} y_{k,j}^m}{n_c} - \frac{\sum_{c=1}^{c} \sum_{y_k^M \in c} y_{k,j}^M}{\sum_{c=1}^{c} n_c}\right)^2
\end{aligned}
$$

然后求导，得：

$$
\begin{aligned}
\frac{\partial\, \mathrm{tr}\left(S_w\right)}{\partial y_{i,j}^M} &= 2\left(y_{i,j}^M - m_{c,j}^M\right) + 2 \sum_{y_i^M \in c} \left(y_{i,j}^M - m_{c,j}^M\right) \cdot \left(-\frac{1}{n_c}\right) \\
&= 2\left(y_{i,j}^M - m_{c,j}^M\right) \\
\frac{\partial\, \mathrm{tr}\left(S_b\right)}{\partial y_{i,j}^M} &= 2n_c\left(m_{c,j}^M - m_j^M\right) \cdot \frac{1}{n_c} + \sum_{c=1}^{c} n_c\left(m_{c,j}^M - m_j^M\right)\left(-\frac{1}{\sum_{c=1}^{c} n_c}\right) \\
&= 2\left(m_{c,j}^M - m_j^M\right)
\end{aligned}
$$

代入，求得损失函数对 $y_{i,j}^M$ 的导数：

$$
\frac{\partial E}{\partial y_{ij}^M} = \left(y_{i,j}^M - d_{i,j}\right) + \gamma\left(y_{i,j}^M + m_j^M - 2m_{c,j}^M\right)
$$

改写为向量形式，其中每一个向量均表示第 i 个样本的对应值：

$$
\frac{\partial E}{\partial \boldsymbol{y}^M} = \left(\boldsymbol{y}^M - \boldsymbol{d}\right) + \gamma\left(\boldsymbol{y}^M + \boldsymbol{m}^M - 2\boldsymbol{m}_c^M\right) \tag{1}
$$

2 反向传播：输出层

输出层仅包含如下仿射变换，其中 $\mathbf{y}^M$ 为输出的预测向量，维度为 M；$\mathbf{x}^N$ 为隐藏层输出向量，维度为 N：

$$
\mathbf{y}^M = W^{M\times N}\mathbf{x}^N + b^M
$$

线性变换过程中权重及偏置项的梯度计算推导如下：

$$
\frac{\partial y_i^M}{\partial w_{ij}} = x_j
$$

$$
\frac{\partial E}{\partial W} = \sum_{i=1}^{N} \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial W} = \frac{\partial E}{\partial y^M}\left(x^N\right)^\top \tag{2}
$$

$$\frac{\partial y_i^M}{\partial x_j^N} = w_{ij}$$

$$\frac{\partial E}{\partial X} = \sum_{i=1}^{N} \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial X} = W^\top \frac{\partial E}{\partial y^M} \tag{3}$$

$$\frac{\partial y_i^M}{\partial b_i^M} = 1$$

$$\frac{\partial E}{\partial b^M} = \frac{\partial E}{\partial y^M} \tag{4}$$

将(1)（反向传播输入）分别代入(2)、(4)，即可得到隐藏层权重和偏置项的梯度。(3)作为上一个隐藏层的反向传播输入继续向前传播。

3 反向传播：隐藏层

隐藏层可视为仿射变换与激活层的组合，若采用 sigmoid 函数，其前向传播可用以下方程描述，其中 $\mathbf{z}^M$ 为隐藏层输出，维度为 M；$\mathbf{x}^N$ 为隐藏层输入，维度为 N：

$$\begin{aligned}\mathbf{z}^M &= sigmoid(\mathbf{y}^M) \\ \mathbf{y}^M &= W^{M\times N}\mathbf{x}^N + b^M\end{aligned}$$

激活层的求导结果如下：

$$\frac{\partial z_i^M}{\partial y_i^M} = z_i^M(1 - z_i^M)$$

即：

$$\frac{\partial \mathbf{z}^M}{\partial \mathbf{y^M}} = \boldsymbol{z}^M \odot (1 - \boldsymbol{z}^M) \tag{5}$$

其中 $\odot$ 表示矢量对位相乘。根据(2)、(4)，隐藏层权重及偏置项梯度计算方法如下：

$$\frac{\partial E}{\partial W} = \sum_{i=1}^{N} \frac{\partial E}{\partial z_i} \frac{\partial z_i}{\partial y_i} \frac{\partial y_i}{\partial W} = \frac{\partial E}{\partial z^M} \odot \frac{\partial z^M}{\partial y^M} \left(x^N\right)^\top \tag{6}$$

$$\frac{\partial E}{\partial X} = \sum_{i=1}^{N} \frac{\partial E}{\partial z_i} \frac{\partial z_i}{\partial y_i} \frac{\partial y_i}{\partial X} = W^\top \left(\frac{\partial E}{\partial z^M} \odot \frac{\partial z^M}{\partial y^M}\right) \tag{7}$$

$$\frac{\partial E}{\partial b^M} = \frac{\partial E}{\partial z^M} \odot \frac{\partial z^M}{\partial y^M} \tag{8}$$

$\frac{\partial E}{\partial z^M}$ 为从后续计算节点反向传播而来的梯度，由(3)（输出层）或(7)（上一隐藏层）反向传播而来。将(5)代入(6)和(8)即得隐藏层权重及偏置项梯度。

4 反向传播：输入层

输入层的权重及偏置项梯度计算公式与隐藏层完全相同，但不再需要对输入 x 求导，直接使用(6)和(8)即可计算。

值得注意的是，一些教材认为输入 x 直接构成输入层。按照这种定义，输入层没有参数，因此也

不需要反向传播。

5 总结

假定输入为 Q 维度向量 x，输入层、隐藏层、输出层维度分别为 P、N、M，则
输出层权重及偏置项梯度：

$$\begin{aligned}\frac{\partial E}{\partial W_3^{M\times N}} &= \left[\left(\boldsymbol{y}^M - \boldsymbol{d}\right) + \gamma\left(\boldsymbol{y}^M + \boldsymbol{m}^M - 2\boldsymbol{m}_c^M\right)\right]\left(\boldsymbol{z}_2^N\right)^\top \\ \frac{\partial E}{\partial b_3^M} &= \left[\left(\boldsymbol{y}^M - \boldsymbol{d}\right) + \gamma\left(\boldsymbol{y}^M + \boldsymbol{m}^M - 2\boldsymbol{m}_c^M\right)\right]\end{aligned}$$

隐藏层权重及偏置项梯度，其中 $\odot$ 表示向量对位相乘。即 $A \odot B = \mathrm{tr}(AB^T)$：

$$\begin{aligned}\frac{\partial E}{\partial W_2^{N\times P}} =& W_3^\top\left[\left(\boldsymbol{y}^M - \boldsymbol{d}\right) + \gamma\left(\boldsymbol{y}^M + \boldsymbol{m}^M - 2\boldsymbol{m}_c^M\right)\right] \odot \boldsymbol{z}_2^N \odot (1 - \boldsymbol{z}_2^N)(\boldsymbol{z}_1^P)^\top \\ \frac{\partial E}{\partial b_2^N} =& W_3^\top\left[\left(\boldsymbol{y}^M - \boldsymbol{d}\right) + \gamma\left(\boldsymbol{y}^M + \boldsymbol{m}^M - 2\boldsymbol{m}_c^M\right)\right] \odot \boldsymbol{z}_2^N \odot (1 - \boldsymbol{z}_2^N)\end{aligned}$$

输入层权重及偏置项梯度：

$$\begin{aligned}\frac{\partial E}{\partial W_1^{P\times Q}} =& W_2^T\left\{W_3^\top\left[\left(\boldsymbol{y}^M - \boldsymbol{d}\right) + \gamma\left(\boldsymbol{y}^M + \boldsymbol{m}^M - 2\boldsymbol{m}_c^M\right)\right] \odot \boldsymbol{z}_2^N \odot (1 - \boldsymbol{z}_1^N)\right\} \odot \boldsymbol{z}_1^P \odot (1 - \boldsymbol{z}_1^P)(\boldsymbol{x}^Q)^\top \\ \frac{\partial E}{\partial b_1^P} =& W_2^T\left\{W_3^\top\left[\left(\boldsymbol{y}^M - \boldsymbol{d}\right) + \gamma\left(\boldsymbol{y}^M + \boldsymbol{m}^M - 2\boldsymbol{m}_c^M\right)\right] \odot \boldsymbol{z}_2^N \odot (1 - \boldsymbol{z}_1^N)\right\} \odot \boldsymbol{z}_1^P \odot (1 - \boldsymbol{z}_1^P)\end{aligned}$$

如果认为输入 x 直接构成输入层，则输入层没有权重和偏置项，用 x 替换 z_1 即可。