2024 春季学期-智能计算系统-作业一

(要求：独立完成，第四周 3 月 20 日前在 BB 平台提交)

1. (10 points)（课本 P39 页习题 2.2）假设有一个只有 1 个隐层的多层感知机，其输入、隐层、输出层的神经元个数分别为 33、512、10，那么这个多层感知机中总共有多少个参数是可以被训练的？

   第一层的权重数量为 $33 \times 512 = 16896$，偏置数量为 512。

   第二层的权重数量为 $512 \times 10 = 5120$，偏置数量为 10。

   所以，总的可训练参数量为 $16896 + 512 + 5120 + 10 = 22538$。

2. (10 points)（课本 P39 页习题 2.3）反向传播中，神经元的梯度是如何计算的？权重是如何更新的？

   首先，计算神经网络输出层的损失函数，据此可以求出损失函数对输出层神经元的偏导数。然后，逐层求出输出层神经元对于中间层神经元的偏导数，并根据链式法则，求出损失函数对中间层神经元的梯度。

   同理计算出损失函数对于每层权重的梯度，设定学习速率，使用梯度下降法更新权重:

$$w = w - \alpha \frac{\partial L(w)}{\partial w} \tag{1}$$

3. (10 points)（课本 P39 页习题 2.5）请简述三种避免过拟合问题的方法。

   - 正则化：包括 L^1 正则化和 L^2 正则化。
   - Bagging 集成方法（或模型平均方法）。
   - Dropout 正则化。
   - 稀疏表示、参数共享、数据集增强、多任务学习、提前终止训练等。

4. (10 points)（课本 P39 页习题 2.6）Sigmoid 激活函数的极限是 0 和 1，请给出它的导数形式并求出其在原点的导数值。

   Sigmoid 函数的表达式为：

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2}$$

   其导数形式推导过程如下：

$$\begin{aligned}
\sigma'(x) &= -\frac{1}{(1+e^{-x})^2} \times (1+e^{-x})' \\
&= -\frac{1}{(1+e^{-x})^2} \times (-e^{-x}) \\
&= \frac{1}{1+e^{-x}} \times \frac{e^{-x}}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}} \times \frac{1+e^{-x}-1}{1+e^{-x}} \\
&= \sigma(x)(1-\sigma(x))
\end{aligned} \tag{3}$$

当 $x=0$ 时，$\sigma(0)=0.5$，所以，$\sigma'(0)=0.25$。

5. (10 points)（课本 P40 页习题 2.9）一种更新权重的方法是引入动量项，即

$$\Delta\omega(n) = \alpha\Delta\omega(n-1) + \alpha^2\Delta\omega(n-2) + \cdots \tag{4}$$

动量项 α 的取值范围通常为 $[0,1]$，这样取值对于权重更新有什么影响？如果取值范围为 $[-1,0]$ 呢？

增加正数动量项后，实现了对权重更新的指数加权平均。越早更新的梯度最终的权重系数越小，而越靠近当前的权重系数越大。使得梯度下降的速度提升，提高了网络收敛的效率，使其能够更优和更稳定的收敛并减少梯度下降过程中的振荡。

增加负数动量项后，由于负系数的幂指数大于其本身，因此会产生与正系数相反的效果。使得权重更新时，产生与上一次变化相反方向的变化，使得网络收敛速度降低、增加了梯度更新过程的振荡。