

人工智能基础算法 第三次作业

2024 年 10 月 9 日

题目内容：

之前我们已经使用了 KNN、线性回归等方法实现了对 MNIST 手写数字数据集的一些特征的识别，本次作业中我们将使用人工神经网络来再次对手写数字的特征进行识别。

我们采用和第一次作业相同的数据集（包含了 30000 个训练样本和 30000 个测试样本）。请使用 Python，Matlab，R 或其他编程语言完成任务并撰写简要报告，提交作业是要包含**所有源码以及报告**，可以补全提供的示例代码，也可以自己另写代码，**本次作业允许调用现有的 python 程序包。**

作业要求：

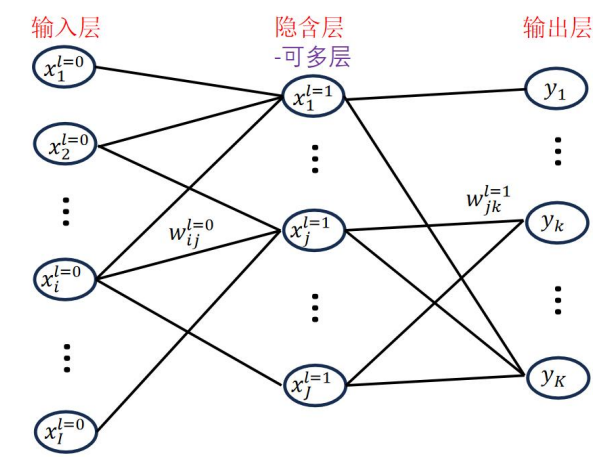
1.（15 分）请推导反向传播算法的更新公式。即：

$$\min_w f(w) = \frac{1}{2} \sum_{k=1}^K (y_k - d_k)^2$$
$$\text{where: } y_k = s(\sum_{j=1}^J x_j^{l=1} w_{jk}^{l=1})$$
$$s(\alpha) = \frac{1}{1 + e^{-\alpha}}$$
$$x_j^{l=1} = s(\sum_{i=1}^I x_i^{l=0} w_{ij}^{l=0})$$

其中 y_k 表示网络的第 k 维输出， $s(\alpha)$ 代表激活函数， $w_{jk}^{l=1}$ 代表隐含层第 j 个节点到输出层第 k 维的权重， $w_{ij}^{l=0}$ 代表输入层的第 i 维到隐含层的第 j 个节点的权重

假设已知 t 时刻的输出值 y_k 和参数值 $w_{ij}^{l=0}, w_{jk}^{l=1}$ ，请给出更新 t+1 时刻，参数 $w_{ij}^{l=0}(t+1), w_{jk}^{l=1}(t+1)$ 的值的计算过程

注：神经网络只由输入层、一个全连接的隐含层和输出层构成（见本次课程 PPT 中 P15 和 P17），请使用 PPT 中给出的符号体系，并给出中间计算偏导数的结果

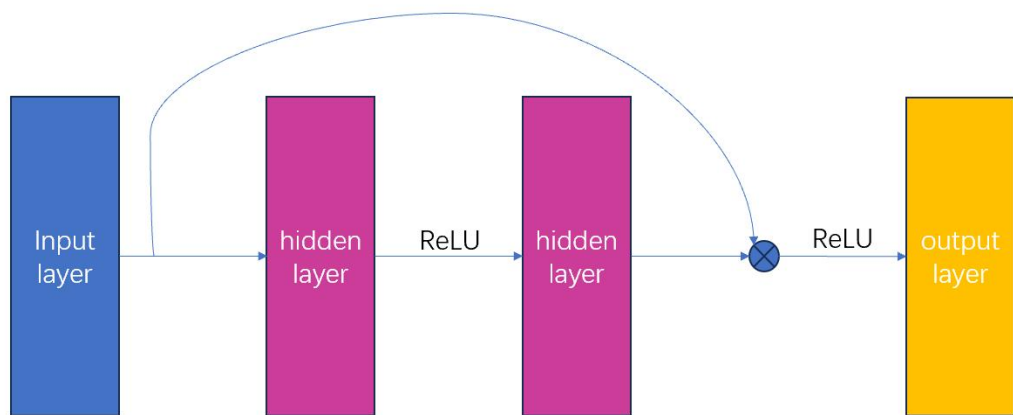


2. (15 分) 设置隐含层中每一层由 100 个节点，分别设置隐含层的层数为 1, 3, 5 层，以 Sigmoid 函数作为激活函数，学习率等参数自行选取，实现对数据集数字的识别,汇报运行时间，绘制训练错误率曲线、测试错误率曲线，并绘制混淆矩阵

3. (15 分) 将激活函数改为 ReLU 函数，重复 2 小题的内容

4. (15 分) 设计隐含层的层数为 1 层，分别设置隐含层变量的个数为 100, 300, 500，以 Sigmoid 作为激活函数，学习率等参数自行选取，实现对数据集数字的识别,汇报运行时间，绘制训练错误率曲线、测试错误率曲线，并绘制混淆矩阵

5. (20 分) 将网络更改为 Resnet（残差神经网络），设置所有的隐含层为 100 个节点，层数为 1，如图构建网络



如果效果不好可自行调整网络中的参数（包括隐含层的层数，变量个数等），实现对数据集数字的识别,汇报运行时间，绘制训练错误率曲线、测试错误率曲线，并绘制混淆矩阵。

6.（20 分）请你根据上面的实验，通过比较它们的结果，讨论隐含层层数、隐含变量的个数等方面对人工神经网络能力的影响，并分析原因

说明：

1. 作业附件 MNIST_dataset 文件夹中提供了数据集，homework.ipynb 文件，供大家参考
2. 请将作业代码和报告打包上传（代码格式可以是.py 文件，也可以是.ipynb 文件），报告中要包含实验结果和对实验结果的说明，打包成.zip 文件
3. 本课程作业禁止抄袭！如有发现，所有作业分数将全部记为 0 分！