|  |  |  |
| --- | --- | --- |
| 第二章 神经网络设计实验 | | |
| 姓名： 薛天钰 | 学号：21371518 | 班级： 212111 |
| **注意：课程实验代码和报告需独立完成，严禁抄袭!如发现有抄袭情况（同年级或者抄袭高年级），相应的实验或报告部分直接0分。** 实验2.1基于三层神经网络实现手写数字分类(60’’) **1.介绍一下“手写数字分类任务”和本实验中使用的数据集。**  MINIST数据集来自美国国家标准与技术研究所（NIST），为手写数字黑白图片和对应标签的集合。  训练集由来自250个不同人手写的数字构成，其中50%是高中学生，50%来自人口普查局的工作人员。测试集也是同样比例，但保证了作者集不相交。  MINIST数据集共有7万张图片，其中6万张是训练集，1万张是测试集。每张图片是28X28的0-9手写数字图片，黑底白字。黑用0表示，白用0-1之间的浮点数表示，越接近1，颜色越白。  图片的标签以一维数组的one-hot编码形式给出，每个元素表示图片对应的数字出现的概率。  MINIST的下载地址<http://yann.lecun.com/exdb/mnist/> ，包含四个部分  ·训练数据集  ·训练数据集标签  ·测试数据集  ·测试数据集标签  **2.本实验的神经网络都包含哪些层？各个层的功能是什么？**  本实验的神经网络包含三个全连接层，前两个全连接层后各跟随一个ReLU激活函数层，第三个全连接层后跟随一个SoftMax损失层。也就是说本实验的神经网络共有6层：    其中，每个全连接层的功能均为：  ·前向传播时 根据当前层神经元激发状态，权重和偏置值计算当前层的输出  ·反向传播时 根据损失函数对当前层输出的偏导计算权重和偏置的梯度以及损失函数对输入的偏导  每个ReLU激活层的功能为：  ·前向传播时计算神经元的激活值  ·反向传播时根据损失函数对输出的偏导计算对输入的偏导  Softmax损失层的功能为：  ·前向传播时计算输出相比标签的损失  ·反向传播时计算损失函数对输入的偏导  **3.网络的训练(training)过程和网络的推理(inference)过程有什么区别？**  训练使用训练集，在不断地迭代中修改参数，接近于模型的建立  推理使用测试集，不修改参数，考察当前模型和参数的适应性，接近于模型的检验  **4.参数和超参数的区别是什么？本实验中都使用了哪些超参数？**  参数是模型内部的可学习变量，通过训练从数据中学习得到。  超参数是用于配置模型和训练过程的参数，需要手动设置。  本次试验中使用的超参数有  学习率  隐藏层数量  隐藏层神经元个数  单次传入的图片数量  训练轮数  **5.什么是正则化？有哪些正则化技术？**  正则化通过添加额外的约束或惩罚项来控制模型的复杂度，防止模型在训练数据上过度拟合。  常见的正则化技术：  L1 正则化（L1 Regularization）：  通过在损失函数中添加权重的绝对值之和来约束模型参数，即添加一个 L1 范数惩罚项。  L2 正则化（L2 Regularization）：  通过在损失函数中添加权重的平方和来约束模型参数，即添加一个 L2 范数惩罚项。  Elastic Net 正则化：  Elastic Net 正则化是 L1 正则化和 L2 正则化的组合。  它在损失函数中同时包含 L1 和 L2 范数惩罚项，可以综合两者的优点，同时减少模型参数并控制参数大小。  Dropout：  通过随机丢弃一定比例的神经元来防止神经网络过拟合，从而提高泛化能力。  Early Stopping：  在训练过程中监测模型在验证集上的性能，并在性能开始下降时停止训练。  数据增强（Data Augmentation）：  通过对训练数据进行随机变换来扩充数据集的技术。  批量标准化（Batch Normalization）：  在神经网络中添加标准化层的技术，可以加速训练过程并减少过拟合的风险。 实验2.2 基于 DLP 平台实现手写数字分类(30’’) **1.什么是量化？量化会带来什么好处/坏处？**  量化：将存储空间较大的数据类型映射到较小的数据类型来表示。  好处：大幅降所需的储存空间，提高计算速度  坏处：损失一些数据，会降低精度。  **2.pycnml、CNML库以及DLP之间的关系是什么？**  DLP是一个深度学习处理器平台，其上已封装好pycnml  Pycnml是深度学习编程库，通过调用DLP上的CNML库中的高性能算子实现了全连接层、卷积层、池化层、ReLU激活层、Softmax损失层等常用的网络层的基本功能，并提供了常用网络层的Python接口。  **3.使用DLP相对于CPU有哪些优势和劣势?**  优势：  DLP专为深度学习设计，具有高优化度的硬件资源，速度较快  DLP能效比较高  DLP有极好的低延迟性，在部分领域至关重要  劣势：  DLP专为深度学习设计，涉及到大量其他运算时CPU更占优势  DLP在推断阶段表现出色，但训练阶段需要大量计算和内存资源，CU更占优。 | | |