

Lab1Part2

一、代码基本结构

对 12 个手写汉字进行分类部分的代码在 `/src/Classifier/` 下的相关文件中：其中 `/src/Classifier/Model.py` 文件定义了 `setup_seed` 方法（即设置 `torch` 中需要用到的随机数种子）、`accuracy_count` 方法（即计算预测值等于真实值的样本个数）、`Model` 类（卷积神经网络模型）；`/src/Classifier/Classifier.py` 包含了训练神经网络的相关代码；`/src/Classifier/AccuracyTest.py` 包含了在测试集上测试准确率的相关代码；此外 `/src/Classifier/model/model.pth` 用以存储模型文件，`/src/Classifier/train_data/` 包含了训练集和测试集。

二、设计实验改进网络及论证

除实验部分外，其余超参数设置包括：迭代次数均为 100 次，网络结构相同（卷积层 1+激活函数+池化层 1+卷积层 2+激活函数+池化层 2+全连接层），以从 12 类汉字样本中各抽样 50 个样本组成的 600 个样本作为测试集，以在该测试集上的准确率作为评判标准。

序号	卷积层 1 输出通道数	卷积层 2 输出通道数	学习率	Batch Size	准确率
1	32	64	0.005	64	98.17%
2	64	128	0.005	64	98.67%
3	128	256	0.005	64	98.83%
4	128	256	0.010	64	99.50%
5	128	256	0.020	64	99.17%
6	128	256	0.010	128	98.50%
7	128	256	0.010	32	99.50%
8	128	256	0.010	16	99.00%

通过序号 1、2、3 的对比实验可以发现，当其他参数不做调整时，在一定范围内，各卷积层输出通道数的增加会导致模型准确率提高，但提高幅度逐渐降低。因为更大的输出通道数意味着有更多的卷积核被应用于卷积过程，模型的容量增大，所以更容易收敛到最优权重，但也会显著提高对性能的消耗。因此，选择了表现较好的 128、256 分别作为卷积层 1 和卷积层 2 的输出通道数。

通过序号 3、4、5 的对比实验可以发现，当其他参数不做调整时，在一定范围内，学习率过大或过小都会导致模型准确率下降。当学习率过小时，该模式需要花费更多的时间收敛到最优权重，因而在有限次（实验中为 100 次）迭代下来不及达到最优情况。而高学习率意味着权重更新的幅度更大，因此可能将花费更少的时间收敛到最优权重，但会导致权重跳动过大，不够准确以致不能达到最优点。因此，选择了较适中的 0.010 作为学习率。

通过序号 4、6、7、8 的对比实验可以发现，当其他参数不做调整时，在一定范围内，Batch Size 过大或过小都会导致模型准确率下降。当 Batch Size 过小时，该模式需要花费更多的时间收敛到最优权重，因而在有限次（实验中为 100 次）迭代下来不及达到最优情况。当 Batch Size 过大时，由于深度学习中随机梯度噪音的 Magnitude 正比于 Learning Rate/Batch Size，随机梯度噪音过小，模型难以达到好的泛化效果，此外也会显著增加对内存的消耗。

因此，选择了较适中的 64 作为 Batch Size。

综上，选择了序号为 4 的一系列参数作为最终应用的参数。

三、对网络设计的理解

在网络设计中，可以采用 Dropout 技术来减少过拟合的发生。过拟合问题即在充分训练后，神经网络的权重与训练样本太过匹配，以至在处理新样本时的表现远不如在训练集上的表现。Dropout 层将丢弃该层中一个随机的激活参数集，即在前向通过中将这此激活参数集设置为 0。该机制能够强制网络变得更加冗余，即该网络将能够为特定的样本提供合适的分类或输出，即使一些激活参数被丢弃。通过 Dropout 机制可以保证神经网络不会对训练样本过于匹配，这将帮助缓解过拟合问题，使模型得到更好的泛化能力。