

实验二思考题

lab1

思考：如果不打乱训练集，会对训练结果产生什么影响？

训练集顺序固定，会导致模型学习会产生过拟合；学习到的东西过分依赖数据的顺序，泛化的能力低

lab2

思考题1：为什么神经网络需要非线性激活函数？如果使用线性激活函数会发生什么？

非线性激活函数能够让神经网络表达非线性的内容，使其能够逼近任意复杂的函数

如果使用线性激活函数，多层神经网络只会等效于单层线性模型，失去强大的表达能力，无法解决非线性的问题

思考题2：观察实验结果，为什么训练准确率会和激活函数选择相关？这与梯度分布有什么关系？

Sigmoid梯度范围在 $(0, 0.25]$ ，当输入值较大或较小时，梯度接近于0，会导致梯度消失问题。

Tanh梯度范围在 $(0, 1]$ 。梯度范围比Sigmoid大，因此它的训练速度是最快的，但在输入值较大或较小时，梯度仍然会接近于0，导致梯度消失问题

ReLU 在正区间梯度恒为1，避免了梯度消失问题，但在负区间梯度为0，可能导致神经元“死亡”。

梯度消失：当激活函数的梯度非常小时，反向传播时梯度会逐层衰减，导致深层网络的权重更新非常缓慢，甚至停止更新。这会使得模型难以训练，尤其是深层网络。

ReLU在正区间的梯度恒为1，避免了梯度消失问题，使得深层网络能够有效训练，并且能够快速收敛，不逊色于tanh

思考题3：ReLU死亡现象的成因是什么？有哪些解决方案？

ReLU激活函数在输入为负时输出为零，且梯度也为零。如果一个神经元的权重更新导致其输入始终为负，那么该神经元的梯度将始终为零，权重将不再更新，这就是死亡现象。

可以调整学习率，改进权重初值，或者使用ReLU的变体函数

思考题4：使用L2正则化后，模型的参数会发生什么变化？为什么这种变化有助于防止过拟合？

L2正则化会惩罚较大的权重值，因此在训练过程中，优化算法会倾向于减小权重的值，权重的分布会变得更加集中，避免出现极端大的权重值。

较小的权重意味着模型的复杂度较低，能够更好地捕捉数据的整体趋势，而不是过度拟合训练数据中的噪声，也会使模型的决策边界更加平滑，避免过于复杂的边界（如锯齿状边界），从而减少过拟合的风险。

思考题5：Dropout为什么能够起到正则化的作用？训练时和测试时的差异处理有什么意义？

每次进行丢弃后，相当于生成了一个新的学习器，这些学习器共享参数，但在结构上有所不同。然后在测试时，相当于为这些学习器做了一个集成，类似集成学习提高了模型的泛化能力。

通过在训练时对保留的神经元进行比例缩放，可以确保训练时和测试时的期望输出一致，从而避免模型在测试时表现异常。在测试时不需要进行随机丢弃，直接使用完整的网络进行计算，简化测试过程。

实验心得与体会

完成本次实验后，自己手写代码深入了解了深度学习的过程，了解了各个函数的意义，还有训练过程中数据的形态变换，对矩阵的形状有了清晰的认知，对python的迭代器有了新的认知。通过实验结果的图像，清晰地了解了每个步骤在实践中的意义。