**2025年春季学期《机器学习》实验报告**

**（第二次实验）**

班级： 230617 学号： 23371007 姓名： 刘嘉明

一、在逻辑回归、SVM 与 MLP 实验中，输入图像在训练前为何需要进行归一化或标准化处理？请分别说明归一化与标准化的原理及其在不同模型训练中的作用。

归一化和标准化处理都是为了把输入数据统一尺度，避免浮动范围大的参数占据主导因素从而影响模型的精确性

归一化: 把数据调整在[0,1]或者[-1,1]的区间,例如Min-Max归一化

标准化: 调整数据的分布,使其均值为0,标准差为1

不论是归一化还是标准化,在上述三种模型的作用都是使输入数据更稳定,统一尺度

在逻辑回顾中,可以避免损失函数变化受大范围数据的影响

在SVM中,可以避免某些特征对距离(margin)影响过大

在MLP中,使各层感知机的输入数据更稳定

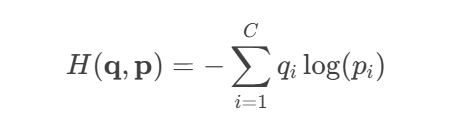
二、请比较逻辑回归、SVM 与 MLP 三种模型在模型结构和表达能力上的差异，并说明为何逻辑回归只能学习线性边界，而MLP可以拟合更复杂的非线性分类问题？

逻辑回归: 通过找到超平面,再通过Sigmoid函数映射到[0,1]范围来表示分类的概率,逻辑回归是一种严格的线性分类器,将数据点一分为二,因此无法处理非线性关系,例如亦或问题

SVM: 根据支持向量计算出超平面,从而找到最优的划分(距离支持向量距离最大的超平面),原本仅支持线性变换,但可以通过核技巧映射到高维空间,进而支持非线性变换

MLP: 由输入层,隐藏层,输出层构成,可以通过多个隐层实现非线性的变换,例如使用非线性激活函数,事实上,单隐藏层的MLP可以逼近任意连续函数

三、在MLP实验中使用了交叉熵损失函数（CrossEntropyLoss）。请结合数学公式说明该损失函数的定义，并解释它为什么适用于多分类任务。相比均方误差（MSE），它在分类任务中具有什么优势？

q是代表真实的分类,p代表预测的分类结果,交叉熵损失函数用于衡量模型预测概率分布和真实概率分布之间的差异

该损失函数的结果反映了将某个样本错分的可能性,最小化交叉熵等价于最大化真实类别的对数似然,符合分类任务的统计假设

该损失函数对错误预测的惩罚更大(log),而MSE对所有误差的惩罚是均等的,不够重视明显的错误预测,不利于后续优化

四、三个实验中都使用了准确率（Accuracy）和混淆矩阵作为评估指标。请说明混淆矩阵中TP、FP、FN、TN的含义，并结合实际图像分类任务举例说明它如何帮助我们识别模型的错误模式。

假设1为Positive,0为Negative

TP: 真实类别为1,预测也为1

FP: 真实类别为0,但预测为1

FN: 真实类别为1,但预测为0

TN: 真实类别为0,预测也为0

混淆矩阵不仅提供了预测的准确率,还具体的解释了错分类型,为优化方向提供了依据

对于B2和F35的分类任务,前者分类为1,后者为0

对于FP: 预测模型将F35错误识别成了B2

对于FN: 预测模型将B2错误识别成了F35

本质上都是没有很好的提取出对于类别的特征,导致分类的超平面不够精准

根据FP和FN的值,可以增加相应的训练样本,例如针对性训练易混淆的图像

1. 在MLP实验中使用了激活函数（如ReLU或Sigmoid）、优化器（如SGD或Adam）、Dropout正则化等策略。请选择其中两项，说明它们在训练深层神经网络中的具体作用和使用时的注意事项。

ReLU激活函数:

作用: 主要有解决梯度消失的问题; 如果是负数入则输出为0,使部分神经元失效,增加稀疏性,提升计算效率; 其线性正区的梯度保持更稳定相比Sigmoid/Tanh，反向传播时权重更新更高效

注意事项: 若神经元权重更新后始终输出负值,则梯度为0,该神经元永久失效; ReLU的输出均值恒为正,可能影响梯度更新方向,可以配合BN层使用

Adam:

作用: 可以结合动量和RMSProp的优点，分别计算梯度的一阶矩和二阶矩,从而动态调整每个参数的学习率,这样以来,每次调整参数的步长就是不固定的,避免步长过短使得寻找最小值点效率过低,也避免因步长过长反复震荡

注意事项: 可能陷入到局部最优解(极小值点); 每次调整参数需要计算一阶矩和二阶矩,计算量显著上升