**2025年春季学期《机器学习》实验报告**

**（第二次实验）**

班级： 230617 学号： 23373179 姓名： 周子强

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。

**（1）归一化和标准化的原理：**

1.归一化：将数据线性缩放到固定范围。保留了原始分布形状，但对异常值敏感。公式为.

2.标准化：将数据调整为均值0、标准差1的分布。不改变数据分布形状，对异常值鲁棒性强，适合非均匀分布的数据。公式为.

**（2）在不同模型中的作用：**

1.逻辑回归Logistic Regression：损失函数和梯度下降优化对特征尺度敏感，若特征尺度差异大，梯度下降在不同方向上的步长不均匀，收敛速度慢。标准化确保了梯度下降优化时各方向步长均衡，同时使正则化惩罚公平作用于所有特征，提升模型泛化能力。

2.支持向量机SVM：核函数基于特征间的距离计算，若特征尺度差异大，大尺度特征会主导距离计算，导致模型忽略小尺度特征的信息，决策边界被少数大尺度特征扭曲，分类性能下降（假设图像像素范围，像素值较大的通道红色通道会主导核函数计算）。

3.多层感知机MLP：激活函数对输入范围敏感，若未标准化，某些神经元可能因为输入值过大而进入激活函数的饱和区（梯度接近0），导致梯度消失。输入特征尺度差异会导致各层权重更新幅度不一致，标准化后反向传播更稳定，加速收敛。

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？

**（1）模型结构差异：**

1.逻辑回归：单层线性模型，由输入特征的线性组合（权重向量与特征的点积）加上偏置项构成，通过sigmoid函数将输出映射到概率（二分类）或softmax函数（多分类）。数学形式为：

其中为sigmoid函数。

2.SVM：线性SVM通过最大化间隔的超平面进行线性分类，决策函数为；非线性SVM利用核函数将输入数据隐式映射到高维特征空间，在高维空间中寻找线性超平面。

3.MLP：由输入层、多个隐藏层和输出层构成，每层包含多个神经元，通过非线性激活函数（ReLU、Sigmoid）连接。数学形式为（单隐藏层）：

其中为隐藏层输出，为激活函数。

**（2）表达能力差异：**

|  |  |  |
| --- | --- | --- |
| **模型** | **表达能力** | **决策边界特性** |
| **逻辑回归** | 仅能表达线性关系 | 线性边界（平面或超平面） |
| **SVM** | 线性核：线性边界；  非线性核：复杂非线性边界 | 依赖于核函数，可高度非线性 |
| **MLP** | 理论上可逼近任意连续函数 | 高度非线性，可拟合任意复杂形状的边界 |

**（3）为什么逻辑回归只能学习线性边界，MLP可以拟合非线性分类问题：**

1.逻辑回归的线性本质：决策函数为输入特征的线性组合，假设两类别的分界满足，这在几何上是一个超平面，无法拟合非线性分布。

2.MLP的非线性能力来源：通过堆叠多个隐藏层，MLP可以逐层提取高阶特征；非线性激活函数引入了非线性变换，使每层的输出不再是输入的线性组合，多层线性激活函数叠加使得模型整体具有非线性表达能力。

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？

**（1）交叉熵损失函数的定义：**

交叉熵损失函数是分类任务中衡量模型预测概率与真实分布之间差异的核心指标。对于多分类问题，其定义如下：

1. Softmax函数归一化：假设模型输出为未归一化的分数，其中为类别数。通过softmax函数将其转换为概率分布：

其中表示样本属于第类的预测概率。

1. 交叉熵损失公式：对于单个样本，真实标签为独热编码的向量（仅正确类别位置为1），交叉熵损失定义为：

实际中，由于真实标签只有一个位置为1（假设第c类），公式简化为：

**（2）交叉熵适用于多分类任务的原因：**

1.交叉熵直接衡量模型预测概率分布与真实分布的差异，通过最小化交叉熵，模型被强制将高概率分配给正确类别，低概率分配给其他类别，符合分类任务的目标。

2.交叉熵与softmax结合时，梯度计算具有简洁形式。对第c类（假设真实类别）的梯度为：

对其他类别的梯度为：

当预测概率，梯度为负，推动梯度增大；对错误类别，梯度为正，推动减小。这种特性使得参数更新方向明确，加速收敛。

**（3）交叉熵相比均方误差MSE的优势：**

1.梯度特性：交叉熵损失梯度与误差成正比，远离最优解时梯度大，加速收敛；均方误差梯度与误差成线性关系，预测接近正确值时梯度趋近于零，易陷入局部最优。

2.概率匹配：交叉熵损失直接优化概率分布的对数似然，强制概率归一化；均方误差最小化预测概率与真实标签的欧氏距离，可能导致概率不归一化。

3.竞争类别：交叉熵损失通过softmax的指数竞争机制，适合处理多分类的互斥性；均方误差无竞争机制，各类别独立优化，可能产生多峰分布。

4.异常值鲁棒性：交叉熵损失对错误预测的惩罚随概率偏离程度指数增长，更敏感；均方误差使用二次惩罚，对严重错误预测的惩罚相对温和。

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。

**（1）混淆矩阵的定义：**

混淆矩阵是分类任务中评估模型性能的核心工具，其核心指标定义如下（以二分类为例）：

1. TP：真正例（True Positive）：真实为A类，模型预测为A类。
2. FP：假正例（False Positive）：真实为B类，模型预测为A类。
3. FN：假负例（False Negative）：真实为A类，模型预测为B类。
4. TN：真负例（True Negative）：真实为B类，模型预测为B类。

**（2）如何利用混淆矩阵：**

1.飞机识别任务中（二分类），混淆矩阵是的，其中对角线是TP和TN，应当值较大。

2.手写数字识别中（多分类），混淆矩阵是的，其中对角线值应当较大。

五、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。

**（1）激活函数（以ReLU为例）：**

1.作用：

a.引入非线性：ReLU通过将线性变换转换为非线性输出，使神经网络能够拟合复杂的非线性函数。

b.缓解梯度消失：在正区间梯度恒为1，避免sigmoid因饱和区导致的梯度消失问题。

c.计算高效：仅需判断输入是否大于0，计算速度快于指数运算的sigmoid。

2.注意事项：

a.神经元死亡：当输入为负时，ReLU输出为0且梯度为0，神经元无法更新权重。

b.输出非零中心化：ReLU输出始终非负，可能导致梯度更新方向偏移，可以通过批量归一化缓解。

**（2）Dropout正则化：**

1.作用：

a.防止过拟合：训练时以概率随即丢弃神经元，迫使网络不依赖单一神经元，增强泛化能力。

b.隐式模型集成：每次迭代采样不同子网络，近似集成了多个模型的预测结果。

2.注意事项：

a.丢弃率选择：输入层建议较小，避免信息损失过大；隐藏层常用0.5，平衡正则化与表达能力；输出层通常不应用Dropout。