

**模式识别与机器学习实验报告**

实验 四

|  |  |
| --- | --- |
| 题 目 | 多层感知机实验 |
| 学 院 | 未来技术学院 |
| 专 业 | 人工智能 |
| 学 号 | 2023112419 |
| 学 生 | 陈铠 |
| 任 课 教 师 | 刘扬 |

哈尔滨工业大学计算机科学与技术学院

2025年秋季

1. **实验内容**

**（一）简要介绍**

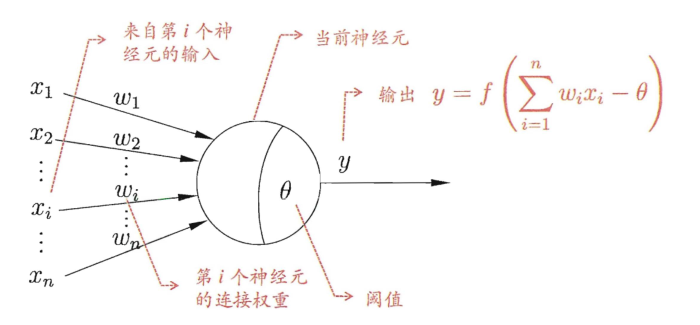
**1.核心任务**

**多层感知机（Multilayer Perceptron, MLP）**是一种基础的前馈神经网络结构，可以实现从输入特征到目标输出的非线性映射。

其核心任务是通过学习一组权重参数，使得网络能够准确地对输入样本进行分类或预测。

**2.数学模型**

多层感知机的基本单元是M-P神经元，其数学表达为：



对于单个隐藏层MLP，设输入层为, 输入层为,隐藏层有m个神经元

则第j个神经元的输出为

第k个输出为

矩阵形式即

其中是输入层到隐藏层的权重矩阵，是隐藏层偏置向量，对向量每个元素独立作用。是隐藏层到输出层的权重矩阵，是输出层偏置向量，是输出激活函数，常使用softmax。

若有多层隐藏层，则

**3.优化方法**

对于分类任务，输出层通常使用softmax激活函数，损失函数采用交叉熵：

反向传播通过链式法则计算梯度：

输出层梯度为

隐藏层梯度为

**（二）实验研究内容**

与线性分类器对比

不同数据量，不同超参数，比较实验效果

1. **实验环境**

google colab云平台

jupyter notebook

1. **实验结果及分析**

本次实验共使用四种数据集，分别为使用高斯分布的类别为3的数据集、sklearn.datasets库中的make\_moons函数生成的数据集、load\_iris鸢尾花数据集、load\_wine葡萄酒数据集。

|  |  |
| --- | --- |
|  |  |
|  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | samples | dimension | categories | test size |
| self | 600 | 2 | 3 | 0.3 |
| make\_moons | 500 | 2 | 2 | 0.3 |
| Iris | 150 | 4 | 3 | 0.3 |
| wines | 178 | 13 | 3 | 0.3 |

**线性分类器和MLP对比**

以下为两种模型的训练损失和验证集准确率，迭代次数均为500

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | lin loss | lin acc | mlp loss | mlp acc |
| self | 0.2375 | 0.878 | 0.2046 | 0.872 |
| make\_moons | 0.3016 | 0.873 | 0.2858 | 0.867 |
| Iris | 0.2567 | 1.000 | 0.1055 | 1.000 |
| wines | 13.2690 | 0.667 | 1.0860 | 0.389 |

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

从表中可以看出，在线性可分的数据集（如Iris）中，两种模型均能取得极高的准确率，说明线性模型已足以完成分类任务。而在具有非线性边界特征的数据集（如make\_moons、自建数据集）上，MLP能通过非线性激活函数（ReLU）捕捉复杂边界特征，表现出更好的拟合能力与灵活性。

但是，在Wine数据集上，MLP的表现不如线性分类器，主要原因可能包括：数据量较小，模型较深时容易过拟合；超参数（学习率、隐藏层大小）未充分调优；特征归一化或输入分布对网络训练的影响较大。总体上，MLP在非线性数据中更具优势，但其性能依赖于合理的网络结构和超参数配置。

**MLP超参数对比**

以下从数据集大小n，隐藏层维度d，学习率lr研究MLP。

迭代次数均为10000次，使用make\_moons

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| n | d | lr | loss | make\_moons acc |
| 500 | 8 | 0.01 | 0.2178 | 0.800 |
| 500 | 8 | 0.03 | 0.0925 | 0.943 |
| 500 | 10 | 0.01 | 0.1713 | 0.886 |
| 500 | 10 | 0.03 | 0.1752 | 0.848 |
| 1000 | 8 | 0.01 | 0.2573 | 0.871 |
| 1000 | 8 | 0.03 | 0.2535 | 0.876 |
| 1000 | 10 | 0.01 | 0.2575 | 0.862 |
| 1000 | 10 | 0.03 | 0.0934 | 0.952 |

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

根据make\_moons数据集的对比结果，学习率(lr)从0.01提升到0.03，模型收敛更快、loss更低、准确率更高 合适的学习率能加速收敛，但过大会引起震荡甚至性能下降隐藏层维度(d)从8到10，准确率变化不大，模型容量略增能提升表示能力，但对小规模任务收益有限数据量(n) 从500增加到1000，模型收敛更稳定、准确率整体提升，更多训练样本可提升泛化性能，验证了数据规模的重要性。

综上，MLP对学习率敏感度最高，其次是数据量，而隐藏层维度在该任务中影响相对较小。合理的超参数选择能显著改善模型性能与收敛速度。

1. **结论**

1.线性分类器 在特征线性可分的任务中具有良好表现，训练简单、收敛快。

2.多层感知机（MLP） 通过引入非线性激活函数与隐藏层结构，能够学习更复杂的决策边界，对非线性任务具有明显优势。

3.MLP 的性能依赖于 数据量、学习率、网络结构等超参数 的综合调优；在小数据集上，若未正则化或调参不当，容易出现过拟合或收敛不稳。

4.实验验证了神经网络模型在非线性分类任务中的有效性，也说明了深度学习中模型复杂度与数据规模的匹配原则。

1. **参考文献**

[1]（美）SHELDON AXLER著；杜现坤，刘大艳，[马晶](https://zhida.zhihu.com/search?content_id=183206416&content_type=Article&match_order=1&q=%E9%A9%AC%E6%99%B6&zhida_source=entity)译. [线性代数应该这样学](https://zhida.zhihu.com/search?content_id=183206416&content_type=Article&match_order=1&q=%E7%BA%BF%E6%80%A7%E4%BB%A3%E6%95%B0%E5%BA%94%E8%AF%A5%E8%BF%99%E6%A0%B7%E5%AD%A6&zhida_source=entity) 第3版[M]. 北京：人民邮电出版社, 2016.10.

[2][周志华](https://zhida.zhihu.com/search?content_id=183206416&content_type=Article&match_order=1&q=%E5%91%A8%E5%BF%97%E5%8D%8E&zhida_source=entity)著. 机器学习[M]. 北京：清华大学出版社, 2016.01.

[3]谢文睿，[秦州](https://zhida.zhihu.com/search?content_id=183206416&content_type=Article&match_order=1&q=%E7%A7%A6%E5%B7%9E&zhida_source=entity)编著. 机器学习公式详解[M]. 北京：[人民邮电出版社](https://zhida.zhihu.com/search?content_id=183206416&content_type=Article&match_order=2&q=%E4%BA%BA%E6%B0%91%E9%82%AE%E7%94%B5%E5%87%BA%E7%89%88%E7%A4%BE&zhida_source=entity), 2021.03.

[4][李航](https://zhida.zhihu.com/search?content_id=183206416&content_type=Article&match_order=1&q=%E6%9D%8E%E8%88%AA&zhida_source=entity)著. 统计学习方法 第2版[M]. 北京：清华大学出版社, 2019.05.