

一.评分标准与论文要求

1.写作质量

- 对应全文语法、语气以及高效表达

2.报告强度

- 行文的连贯性
- 是否有效地反驳了假论文

3.有效实验设计

- 设计健壮、有效的实验
- 通过实验有效体现反驳

4.结果讨论

- 实验数据分析
- 实验数据可视化

5.结果质量

- 实验数据质量

6.理论分析

- 数学与理论推导

7.可持续性分析

- 对自然环境的影响

8.同行评审

- 互相评价工作质量

9.多模型实验

- 采用多种模型实验

10.大特征数性能

- $n \geq 20$ 时的性能是否满足要求

11.强理论分析

- 理论分析为何原论文不可行

12.额外可持续性分析

- 没看懂

13.模型自信度分析

原文未来工作

1.不同的机器学习模型

- 在方法论体现

2.不同的学习器

- 在实验设置分析

3.收敛分析

- 在收敛性实验体现

4.函数近似

- 在MLP分析体现

5.预测不确定性

- 在自信度实验体现
-

二.论文结构

1.题目

2.摘要

- **背景**: 氮石-2.0论文提出机器学习算法无法在该数据集上成功实现二分类任务, 以及认为通用近似定理无效
- **观点**: 我们认为通用近似定理是正确的, 假论文的问题只是“模型-问题不匹配”+“不充分的超参数调优”
- **工作**: 我们通过理论分析和实验设计反驳了假论文的观点, 以及讨论了原论文的未来工作
- **实验**: 我们的实验取得了什么性能效果, 以及还做了什么其他的实验

3.引言

- **背景**: 讨论假论文的观点
- **假设**: 提出我们的观点
- **结构**: 介绍本文的行文逻辑
- **贡献**: 通过理论和实践证明假论文的观点错误, 以及提出了什么新的东西

4.方法论

- **基线模型**: 证明多项式逻辑回归的局限性以及他为什么在这个数据集上存在问题
- **MLP模型**: 理论证明可行, 结合通用近似定理论证为何可行
- **SVM(RBF核)模型**:
- **XGBoost模型**:
- **学习器**: 比较SGD与其他学习器, 如Adam的差别

5.实验设计

- **实验环境介绍**: CPU、GPU、Python以及库版本等
- **数据集介绍/分析**
- **超参数设置/网格搜索**
- **性能实验**: 与基线和目标准确率比较
- **收敛实验**: Loss曲线

- 自信度实验：校准分析

6.结果分析

- 性能实验分析
- 收敛实验分析
- 自信度实验分析
- 计算功耗/环境可持续性分析（量化计算成本）

7.讨论

- 是否出色的完成了任务
- 对原文未来工作的回应
- 本论文的局限性