反驳思路

1.他的观点

论文的明确目标是“对围绕现代机器学习模型的炒作提出质疑” 并“突显关键局限性“

2.他的挑战方向

1. 理论挑战：挑战函数近似定理
2. 实例模型挑战：挑战LLM能力

3.他的证明方式

1. 基于多项式基扩展的对数几率回归-->挑战函数近似定理
2. 使用GPT-2进行特征处理的对数几率回归-->挑战LLM能力

4.他的未来工作

- 不同的机器学习模型
- 不同的学习器
- 收敛性分析
- 函数近似
- 预测不确定性

我的反驳逻辑

1. 证明他的证明方法是存在问题的----提出猜想

1.1 基于多项式基扩展的对数逻辑回归： 基于多项式基扩展的对数逻辑回归是通用近似器，但效果不好不是由于通用近似定理失效

根据 Stone-Weierstrass Theorem，多项式基扩展函数满足以下三个条件

1. 是一个代数
2. 能够分离点
3. 处处非零

所以多项式函数是一个通用近似器，可以拟合任意的函数，在这里是决策边界 (前提是数据集满足Stone-Weierstrass定理的限定条件)

他的问题来源于参数选择和维度灾难

1.多项式的次数选择：定理只保证存在，但是没有给出求多项式次数的方法

2.维度灾难：多项式扩展会导致特征数急剧增加，导致严重过拟合

1.2 基于GPT-2的对数逻辑回归： GPT-2是文本模型，低准确率来自架构选择错误而不是大模型的能力

1.模型与声明不匹配

- 文中说愚弄了GPT-3，实验使用GPT-2

2.实验方法严重受限

- 逻辑回归是一个线性分类器。这个实验的失败**仅仅证明了 Kryptonite-2.0 在 GPT-2 的默认嵌入空间中不是线性可分的。**这根本没有测试 LLM 学习解决问题的能力，只测试了它"开箱即用"的表示能力

3.输入方法不自然且有损

- 作者将数值向量转换为**文本字符串**（例如 `"0.1, 0.5, -0.2..."`）并将其作为"提示"
- LLM（尤其是 GPT-2）是为处理*自然语言*而训练的，而不是处理数值向量的文本表示。
- 这种输入方式本身就是有损的。作者自己在结论中也无意中承认了这一点，他们说这个过程"**破坏了...有用的信息**" 。失败的不是 GPT 模型，而是作者选择的、糟糕的"提示工程"。

4.对工具的根本性误用

- 作者将 LLM 视为与多项式扩展类似的*固定*特征提取器 ， 这是对现代神经网络的误解。

2. 证明我的猜想

证明多项式基扩展对数几率回归效果不好的原因不是通用近似定理失效

2.1 维度灾难

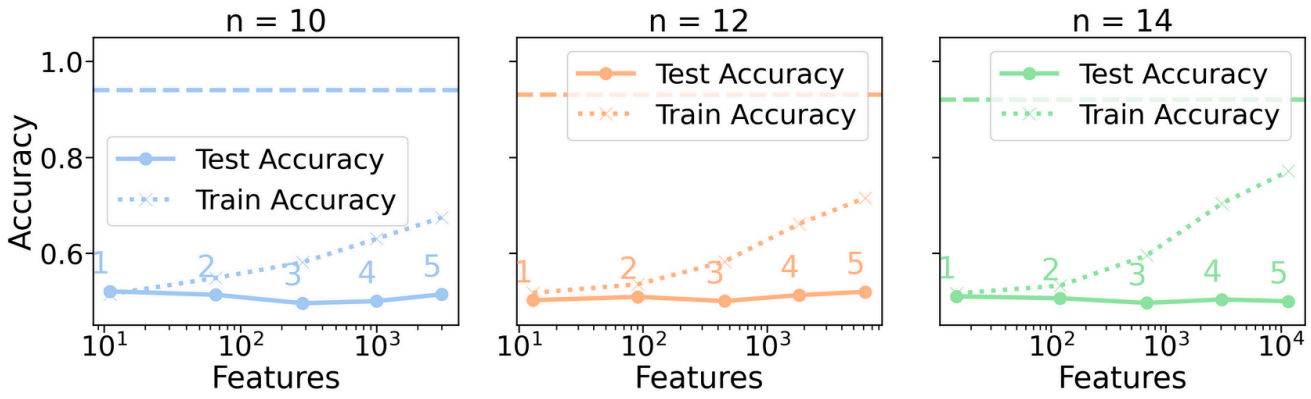
扩展后特征数量计算公式为（包含偏置/截距项）：

$$D = \frac{(n+k)!}{n!k!}$$

其中，

- n为原始特征个数
- k为多项式扩展次数
- D为扩展后的新特征维度

由于特征数量增长快过样本数量增长，模型由于维度灾难会导致非常严重的**过拟合**



2.2 多项式参数选择

使用采用RBF核的SVM证明，多项式次数选择是导致文中模型出问题的因素

RBF核的公式如下：

$$K_{RBF}(x,z)=\exp(-\gamma||x-z||^2)$$

齐次多项式核的公式如下：

$$K_{poly,k}^{homo}(x,z)=(<x,z>)^k$$

展开RBF核

$$||x-z||^2=<x-z,x-z>=<x,x>-2<x,z>+<z,z>=||x||^2-2<x,z>+||z||^2$$

代回核函数

$$K_{RBF}(x,z)=\exp(-\gamma(||x||^2+||z||^2))\cdot\exp(2\gamma<x,z>)$$

对最后一项进行泰勒展开

$$\exp(2\gamma<x,z>)=\sum_{k=0}^{\infty}\frac{(2\gamma)^k}{k!}(<x,z>)^k$$

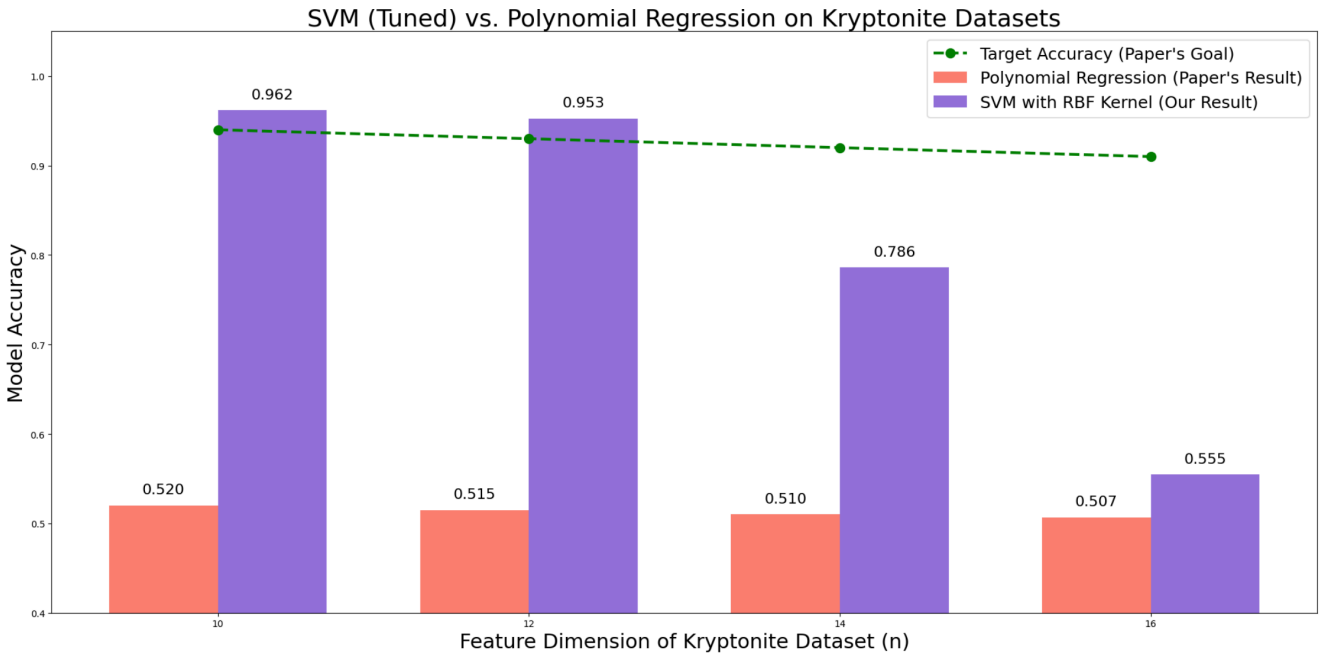
将这个无限和代回RBF核的表达式

$$K_{RBF}(x,z) = \exp(-\gamma(||x||^2 + ||z||^2)) \cdot \sum_{k=0}^{\infty} \frac{(2\gamma)^k}{k!}(< x, z >)^k$$

RBF核可以看作一个从k=0到k=∞的所有齐次多项式核的加权线性组合，所以使用SVM(RBF)等效为测试和使用了所有可能的多项式次数的特征

使用SVM(RBF)模型在数据集的特征数较小的时候性能较好，证明了多项式对数逻辑回归的性能不足来源于多项式参数选择问题

SVM(RBF)的核基于距离计算，在高维度数据性能下降来源于高维数据维度灾难的距离度量失效问题



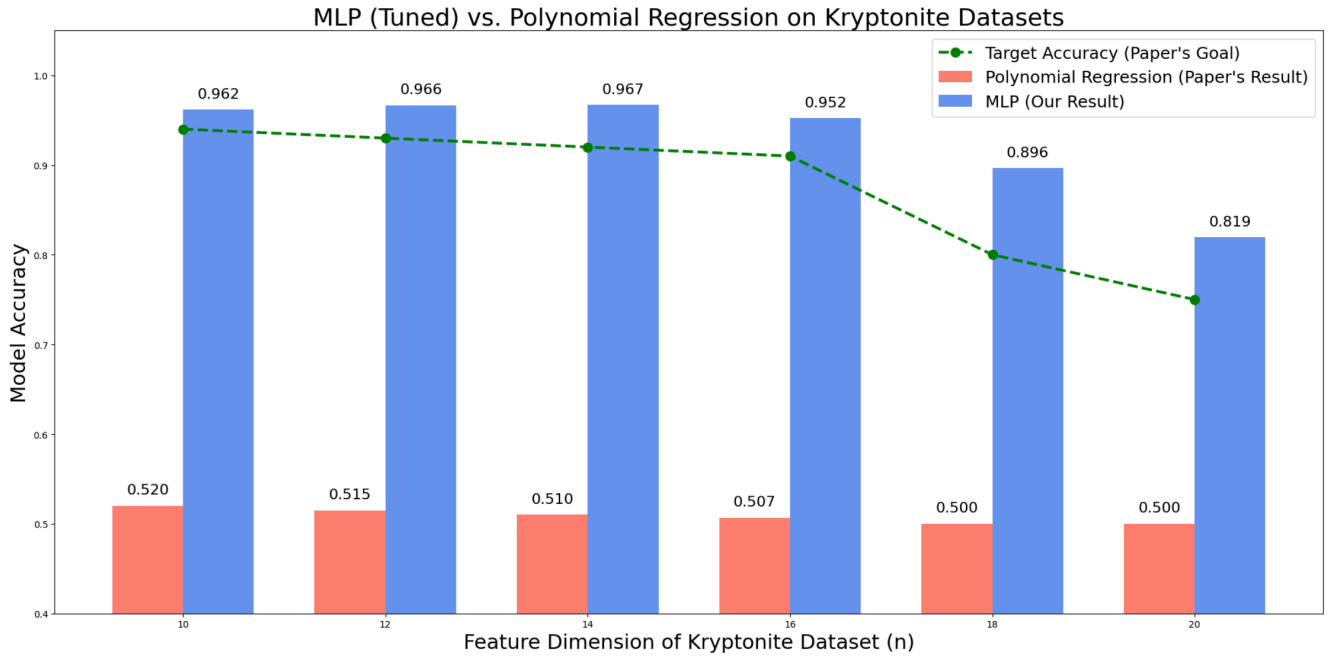
2.3 寻找有效的模型

单隐藏层MLP

根据Cybenko提出的定理，一个使用 S 形激活函数的 MLP 是一个通用近似器

$$F(x) = \sum_{i=1}^{N} w_i \sigma(w_i^T x + b_i)$$

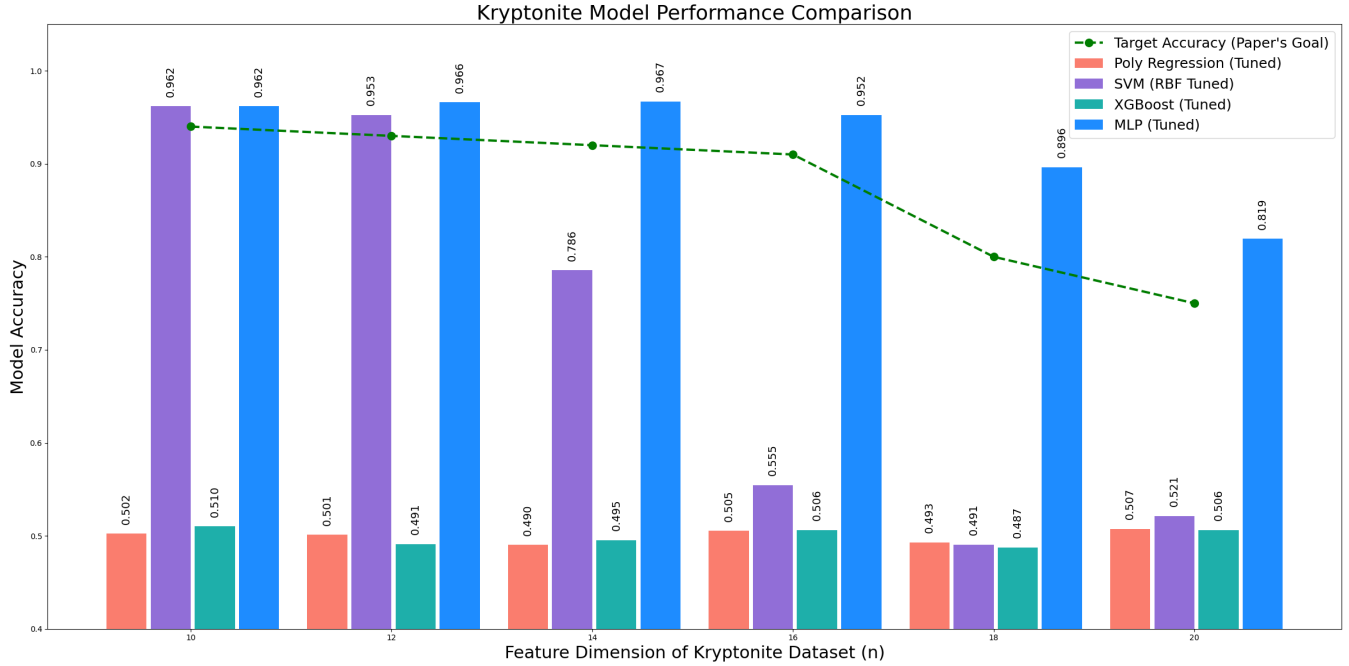
MLP **自动执行了特征选择和重加权**，它学会忽略高维噪声，所以在高维度表现也很好



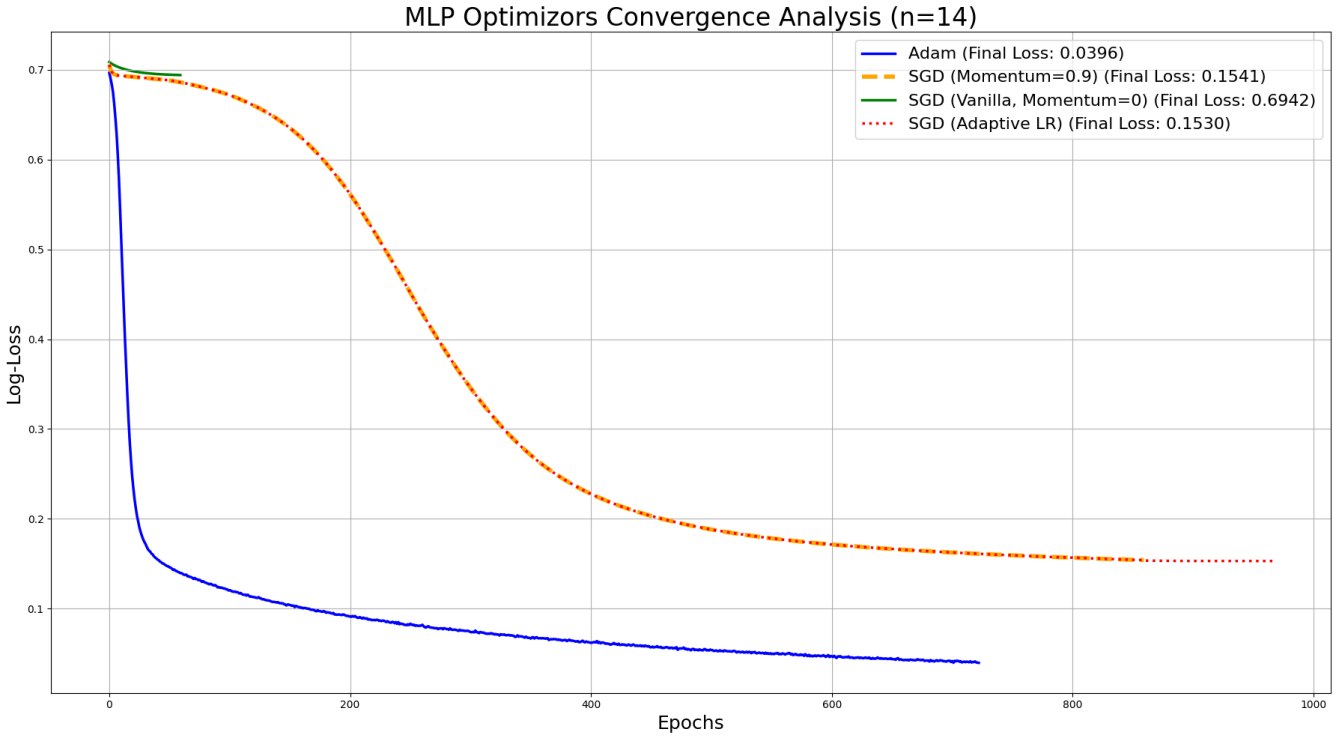
证明基于GPT-2的对数逻辑回归失效的原因是：GPT-2是文本模型，低准确率来自架构选择错误而不是大模型的能力

3. 未来工作回应

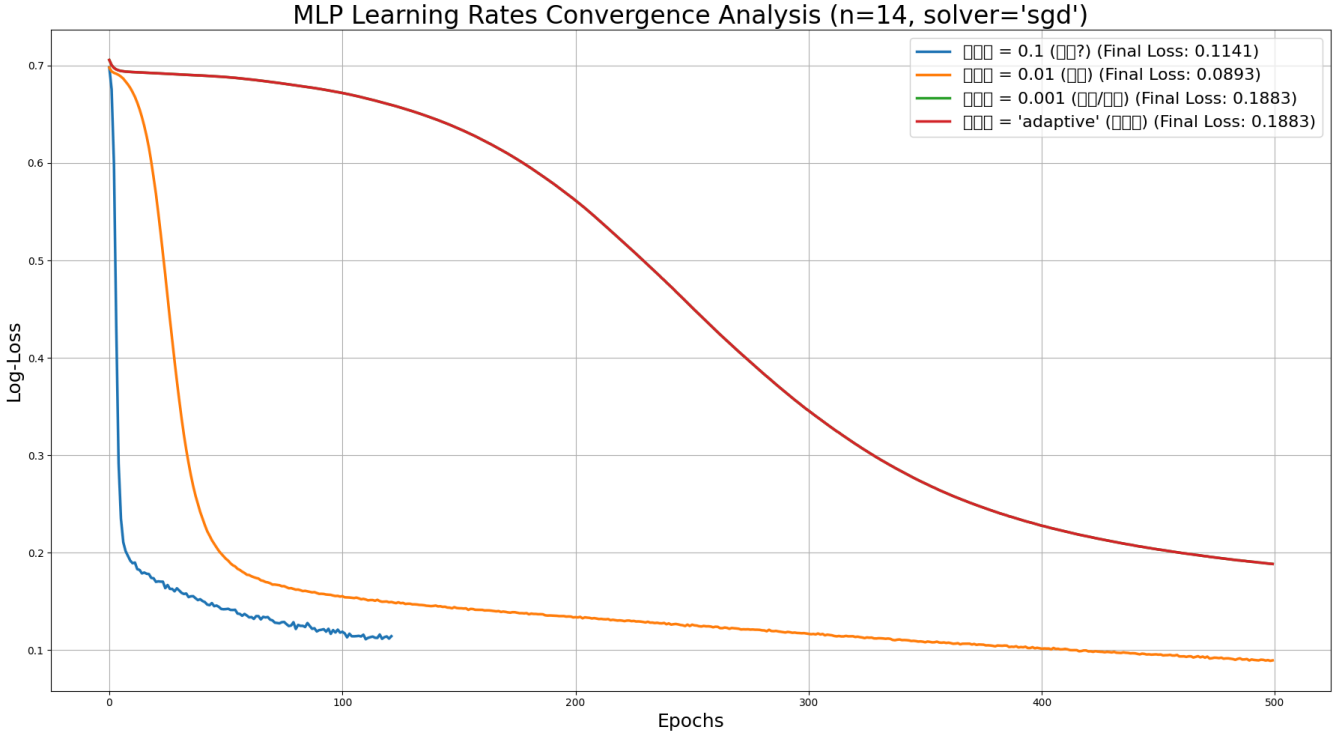
3.1 不同模型



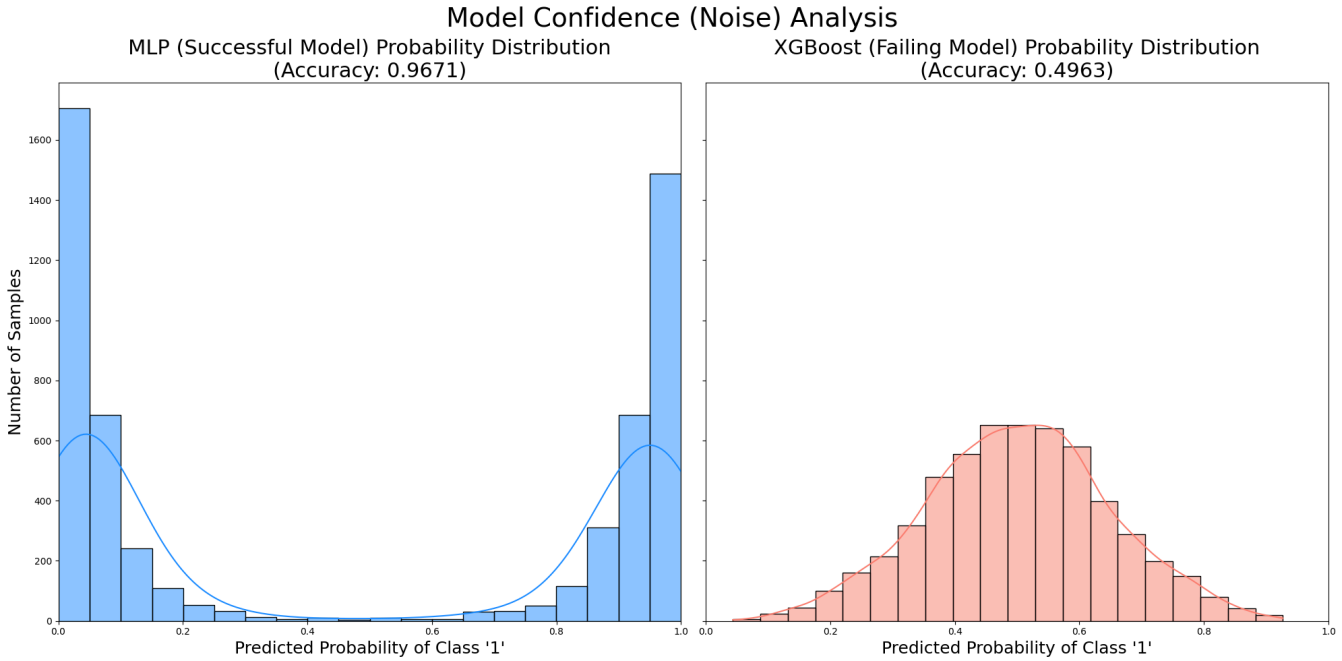
3.2 不同学习器



3.3 收敛性分析（学习率）



3.4 自信程度分析


Model Confidence (Noise) Analysis
MLP (Successful Model) Probability Distribution
(Accuracy: 0.9671)
XGBoost (Failing Model) Probability Distribution
(Accuracy: 0.4963)
Number of Samples
Predicted Probability of Class '1'
Predicted Probability of Class '1'