对带擦除的非交互式相关性蒸馏（NICD）中多数函数最优性的反例

第0部分：引言与背景设定

0.1 标题、作者及单位

【原文翻译】

对带擦除的非交互式相关性蒸馏（NICD）中多数函数最优性的反例

ChatGPT*[1]，Paata Ivanisvili[2]，以及 Xinyuan Xie[2]

[1]OpenAI，美国加利福尼亚州旧金山，邮编 94110，chat.openai.com [2]加州大学欧文分校数学系，美国加利福尼亚州，邮编 92697，{pivanisv,xinyuax7}@uci.edu

【专业解读】

这篇论文的标题信息量巨大，我们可以像剥洋葱一样一层层解析。首先，“反例”是数学研究中的一个“大杀器”。在数学中，要证明一个命题普遍成立，你需要进行严谨的逻辑推导；但要推翻一个命题（或猜想），你只需要找到一个不符合该命题的具体例子，这个例子就是“反例” 。这篇论文的核心贡献，就是找到了这样一个反例。

其次，这个反例是针对什么问题的呢？是关于“多数函数最优性”的。这里的“多数函数”就像一个民主投票系统：输入一堆“赞成”（+1）或“反对”（-1）的意见，如果赞成的多，结果就是“赞成”，反之亦然。在某个特定的问题背景下，人们长期以来认为这个“多数投票”机制是表现最好的，即“最优的”。而这个特定的问题背景，就是“带擦除的非交互式相关性蒸馏（NICD）”。这个术语听起来很复杂，但我们可以把它想象成一个通信难题：两个间谍（比如爱丽丝和鲍勃）收到了一份加密电报的两个副本，但由于信号干扰，两个副本都丢失了一些信息（这就是“擦除”）。他们需要在不互相联系（“非交互式”）的情况下，各自对残缺的电报进行解码，并得到一个完全相同的密钥比特（这就是“相关性蒸馏”）。这篇论文就是要证明，在解决这个通信难题时，最简单的“多数投票”解码方式并非总是最优的。

最引人注目的一点是作者列表。除了两位人类数学家，第一作者赫然是“ChatGPT”。这不仅仅是为了新奇，它标志着科学研究范式的一次深刻变革。人工智能（AI）不再仅仅是用于数据处理和计算的工具，而是开始成为能够提出新颖见解、甚至在悬而未决的数学问题上取得突破的“合作伙伴” 。这篇论文本身，就是AI辅助科学发现的一个里程碑式案例 。这种人机协作模式——AI负责在巨大的可能性空间中进行广度优先的探索和搜索，筛选出有潜力的候选方案，而人类专家则负责进行深度优先的严格验证、理论解释和背景构建——预示着未来科学研究的一种强大新模式。AI的“蛮力”搜索能力弥补了人类直觉的局限，而人类的深刻洞察力则赋予了AI发现的意义。

0.2 摘要与前言

【原文翻译】

摘要 我们要求 GPT-5 Pro 在一个公开的开放问题列表（西蒙斯基金会的“计算机科学中的实分析”合集）中寻找反例。经过几次数值实验，它为带擦除的非交互式相关性蒸馏（NICD）问题提出了一个反例：即一个5比特的布尔函数，当擦除参数 $p = 0.40$ 时，其 $E|f(z)|$ 的值严格大于5比特多数函数的值。在这篇非常简短的笔记中，我们记录了这一发现，精确地陈述了问题，给出了这个显式函数，并用手一步步地验证了计算过程，以便在没有计算机的情况下也能进行核对。此外，我们证明了对于每个固定的奇数 n，在 $p = 0$ 的一个邻域内，多数函数（在无偏布尔函数中）是最优的。我们将此视为AI在理论计算机科学领域贡献的一个“小火花”：虽然现代大语言模型（LLMs）通常在文献和数值方面提供帮助，但在这里，一个具体的有限反例出现了。

【专业解读】

摘要是论文的浓缩精华，它告诉我们三件核心事情。第一，研究的起点非常“高”：研究者们直接让AI去挑战一个领域内公认的难题列表——西蒙斯基金会的这个合集，可以看作是理论计算机科学领域的“难题榜”，挑战上面的问题本身就极具价值。第二，AI的贡献非常具体和关键：它不是模糊地指点方向，而是直接给出了一个精确的数学对象（一个5比特布尔函数）和一个具体的场景（当信息丢失率为40%时），并声称在这个场景下，这个新函数比传统的“多数函数”表现更好。第三，人类科学家的角色不可或缺：他们没有盲目相信AI的结果，而是进行了最严格的“手动验算”，确保结果的绝对可靠，并且进一步从理论上分析了这个反例的适用范围，证明了在信息丢失率极低时，“多数函数”仍然是王者。

作者们用“小火花”这个词来形容AI的贡献，这个比喻非常精妙。它既承认了AI的创造力——能够在一个人类探索了十多年的问题上发现新的东西，又保持了科学的审慎——这并非通用人工智能的降临，而是在一个特定问题上，通过强大的计算和模式匹配能力实现的突破。这颗“火花”的意义在于，它点燃了一种新的可能性：AI可以成为数学家和理论科学家的“灵感催化剂”。

【原文翻译】

简短前言 一个AI为一个存在了十多年的开放问题提出一个可行的反例，据我们所知，这是首个被公开报道的成功案例。在我们测试的大多数问题中，模型除了提供文献指引和进行小规模数值计算外，贡献甚微。相比之下，这里的结果是一次明确的胜利；如果有人事先向我们展示这个例子，我们会印象深刻。这个结果也呼应了与线性阈值函数（LTFs）相关的“多数最不稳定”猜想的经典反例：即使AI仅仅是将一个已知的LTF反例模式应用到一个新的情境中并进行了验证，那也仍然值得称赞。

【专业解读】

这段前言充满了兴奋之情，作者们毫不掩饰地将这一发现称为一次“明确的胜利”。这背后的潜台词是，在AI辅助科研的道路上，有无数次的失败和尝试，AI常常只能做一些文献整理、代码编写等辅助性工作。而这一次，AI真正触及了科研的核心——发现与洞见。

作者还提到了一个重要的类比：与“多数最不稳定”猜想的反例的相似性。这不仅是为了说明这个反例的结构并不孤立，更是暗示了AI可能的工作方式。AI可能不是从第一性原理出发进行“思考”，而是通过学习海量的数学文献和案例，识别出一种被称为“两重三轻”（two heavy + three light）的有效模式，并“猜测”这种模式可能也适用于当前这个看似不同但结构相关的问题。这是一种强大的“类比推理”能力，尽管是基于数据和结构的，而非人类的抽象理解。这个过程本身就极具启发性：AI通过其强大的模式识别能力，可能可以发现不同数学领域之间隐藏的结构性关联，而这些关联可能会被人类研究者忽略。这正是AI作为研究伙伴的巨大潜力所在 。

第1部分：数学基础的铺垫

1.1 问题陈述 - 布尔函数与超立方体

【原文翻译】

1 问题陈述 我们在配备了均匀概率测度的超立方体 $\{-1,1\}^n$ 上进行研究。一个布尔函数是 f : $\{-1,1\}^n \to \{-1,1\}$。如果 $E[f] = 0$，我们称 f 是无偏的。一个充分条件是奇性：$f(-x) = -f(x)$。

【专业解读】

这一段为整个问题搭建了数学舞台。首先，什么是“布尔函数”？简单来说，它就是一个处理“是/非”逻辑的函数 。它的输入是一组二进制信号（比如计算机中的0和1），输出也是一个二进制信号 。在数字电路中，每一个逻辑门（AND, OR, NOT）都对应一个布尔函数 。为了数学处理上的方便，论文没有用 $\{0,\ 1\}$ ，而是用了 $\{-1,\ 1\}$ 来表示两个逻辑状态。这样做的好处是，计算期望（平均值）时，一个结果为0的期望值 $E[f]=0$ 就直观地表示两种输出结果 $+1$ 和 -1 出现的概率完全相等，不多不少。

这样的函数被称为“无偏的”（unbiased），就像一枚公平的硬币，正反面出现的概率都是一半。一个很容易满足“无偏”条件的特性是“奇性”（oddness），即 $f(-x) = -f(x)$。这在几何上非常直观：想象一个投票系统，如果所有人都投了反对票（所有输入从 x 变成 $-x$），那么最终的选举结果也应该正好相反（输出从 $f(x)$ 变成 $-f(x)$），这体现了一种对称性和公平性。

而“超立方体” $\{-1,1\}^n$ 则是所有可能输入构成的空间。当 $n = 1$ 时，它就是两个点 -1 和 1 。当 $n = 2$ 时，它就是平面上的四个点 $(-1,-1)$ ，$(-1,1)$ ，$(1,-1)$ ，$(1,1)$ ，构成一个正方形。当 $n = 3$ 时，它就是三维空间中的八个顶点，构成一个立方体。对于更高维度的 n，

我们称之为“超立方体”。布尔函数 f 的作用，就是给这个超立方体的每一个顶点，都贴上一个 $+1$ 或 -1 的标签。

1.2 问题陈述 - 多线性扩展

【原文翻译】

多线性扩展。每个 f 都有一个唯一的多线性多项式表示 $f(z) = \sum_{S\subseteq[n]} \hat{f}(S) \prod_{i\in S} z_i, z \in [-1,1]^n$; 其中 $\hat{f}(S) = E$，且 $x \sim \mathrm{unif}(\{-1,1\}^n)$。

【专业解读】

布尔函数原本只在超立方体的顶点（即输入分量全是 $+1$ 或 -1 的地方）有定义。但我们之前提到的“擦除”模型，会产生 0 这样的输入值。那么，当输入是 $(1, 0, -1)$ 这样的“非顶点”时，函数值是多少呢？“多线性扩展”就是解决这个问题的关键。它通过一个数学公式，将函数的定义域从离散的顶点扩展到了整个实心超立方体（即输入分量可以是 -1 到 1 之间的任意实数）。

这个过程好比，我们只知道一个房间八个角落的温度，现在想要构建一个模型来预测房间内部任意一点的温度。多线性扩展就是这样一种“插值”模型。它给出的公式 $f(z) = \sum_{S\subseteq[n]} \hat{f}(S) \prod_{i\in S} z_i$ 是这个函数在整个空间中的“温度分布图”。

这个公式在布尔函数分析领域被称为“傅里叶展开” 。其中的系数 $\hat{f}(S)$ 称为傅里叶系数，它们有着非常重要的物理意义。可以把它们理解为不同输入组合的“影响力权重” 。例如：

- $\hat{f}(\emptyset)$ 是函数的平均值（对于无偏函数，它为0）。
- $\hat{f}(\{i\})$ 衡量的是第 i 个输入变量单独对输出结果的影响力有多大。
- $\hat{f}(\{i,j\})$ 衡量的是第 i 个和第 j 个输入变量的“协同作用”对输出的影响力。 这个多线性扩展不仅解决了输入为0的问题，还把一个组合逻辑问题转化为了一个代数问题，让我们可以使用微积分等分析工具来研究它。

1.3 问题陈述 - 擦除模型与目标

【原文翻译】

擦除模型。固定 $p \in (0,1)$。令 $z \in \{-1,0,1\}^n$ 的坐标是独立的，满足 $\Pr[z_i = 1] = \Pr[z_i = -1] = \frac{p}{2}, \Pr[z_i = 0] = 1 - p$ 这是标准的二进制擦除NICD模型（参见 Yang ）。

目标。对于无偏布尔函数 f，定义 $\Phi_p(f) := E_z \mid f(z) \mid$， 其中 $f(z)$ 是在擦除样本 z 处求值的多线性扩展。

开放问题（记录在西蒙斯列表中）。对于 $p < \frac{1}{2}$，$\Phi_p(f)$ 是否由多数函数（在无偏 f 中）最大化？（当 $p \geq \frac{1}{2}$ 时，独裁函数是最优的；参见 。）我们参考西蒙斯合集以获得简明的陈述。

【专业解读】

这里正式定义了问题的核心。首先是“擦除模型”，它精确描述了信息是如何丢失的。想象一下，原始信息是一个由 $+1$ 和 -1 组成的比特串。在传输过程中，每个比特有 $1-p$ 的概率被“擦除”掉，变成 0 。如果一个比特没有被擦除（这发生的概率是 p)，那么它有一半的概率是 1 ，一半的概率是 -1 。这个 p 值就是“信息保真度”，p 越大，信息保留得越完整；p 越小，信息丢失得越严重。

接下来是“目标函数” $\Phi_p(f) := E_z \mid f(z) \mid$ 。这正是对我们之前提到的“通信难题”的数学化。在收到了带有擦除（0值）的输入串 z 后，我们用多线性扩展计算出函数值 $f(z)$。这个值可能是一个介于-1和1之间的小数。它的绝对值 $\mid f(z) \mid$ 可以被理解为我们对输出结果的“信心”或“确定性”。如果 $\mid f(z) \mid$ 接近1，说明我们非常确定结果是 $+1$ 或 -1 。如果 $\mid f(z) \mid$ 接近0，说明输入信息丢失得太严重，我们几乎无法做出判断。因此，最大化 $\Phi_p(f)$，即最大化 $\mid f(z) \mid$ 的期望值，就等同于寻找一个在信息大量丢失的情况下，平均而言能做出最可靠、最自信判断的解码函数 f。

最后，“开放问题”被清晰地提了出来：在过去的十多年里，研究者们普遍猜测，当信息丢失不算太严重（$p < 1/2$）时，最稳健、最自信的函数就是我们最熟悉的“多数函数”（Majority function）。这个猜想直观上很有道理，因为它对所有输入一视同仁，看起来最能抵抗随机的擦除。然而，这篇论文将要证明，这个直觉是错误的。

第2部分：反例及其背景

2.1 p = 0.40 时的有限反例

【原文翻译】

2 p = 0.40 时的有限反例 设 $n = 5$ 且 $p = \frac{2}{5} = 0.40$。考虑线性阈值函数 $f(x_1, ..., x_5) = \operatorname{sgn}(x_1 - 3x_2 + x_3 - x_4 + 3x_5)$（约定当 $t > 0$ 时 $\operatorname{sgn}(t) = 1$，当 $t < 0$ 时为-1；注意在我们的情况下 $t = 0$ 不会发生）。那么 $f(-x) = -f(x)$，因此 f 是无偏的。

定理 1。对于 $p = 0.40$，我们有 $\Phi_{0.40}(f) = \frac{2689}{6250} = 0.43024$，$\Phi_{0.40}(\mathrm{Maj}_5) = \frac{5363}{12500} = 0.42904$ 因此 $\Phi_{0.40}(f) > \Phi_{0.40}(\mathrm{Maj}_5)$。特别地，对于 $n = 5$，在 $p = 0.4$ 时多数函数已经不是最优的了。*证明*。见附录A。

【专业解读】

这里，论文抛出了核心的“炸弹”——那个传说中的反例。让我们来仔细审视这个“挑战者”函数：$f(x_1, ..., x_5) = \mathrm{sgn}(x_1 - 3x_2 + x_3 - x_4 + 3x_5)$。

这个函数属于一类被称为“线性阈值函数”（Linear Threshold Function, LTF）的函数 。你可以把它想象成一个“加权投票”系统。在一个标准的5人多数投票（ Maj5 ）中，每个投票者（输入 x_i）的权重都是1。而在这个反例函数中，投票者的权重被区别对待了：第2和第5个投票者是“超级投票者”，他们的权重高达3（其中 x_2 的权重是-3，意味着它投的是反对票），而第1、3、4个投票者则是普通投票者，权重为1。这种“权重不对称”的设计是它成功的关键。

为什么这种不对称的设计会更优越？直观地想，当信息丢失率 $p = 0.4$（意味着有60%的比特被擦除）时，情况已经相当糟糕了。标准的多数函数依赖于收集到足够多的“幸存”比特来做出判断，它是一种“民主”但分散的策略。而这个反例函数则采取了一种“精英”或“押宝”的策略：它极度依赖于第2和第5个比特。它赌的是，在这两个“关键先生”中，至少有一个能幸存下来（不被擦除）。只要其中一个幸存，就能对最终结果产生巨大的影响。在 $p = 0.4$ 这个特定的噪声水平下，这种“押宝”策略的平均收益，恰好以微弱的优势超过了“民主”策略。

下面的表格直观地展示了这场对决的结果，它将论文的核心发现浓缩成一次清晰的数值比较。

表1：在 p=0.40 时函数稳健性对比

对比函数	公式	$\Phi_{0.40}$ 值 (稳健性)	计算来
反例函数 f	$\mathrm{sgn}(x_1 - 3x_2 + x_3 - x_4 + 3x_5)$	**0.43024**	$\frac{2689}{6250}$
5比特多数函数 Maj5	$\mathrm{sgn}(x_1 + x_2 + x_3 + x_4 + x_5)$	0.42904	$\frac{5363}{12500}$

这个表格无可辩驳地证明了论文的核心论点：在 $n = 5, p = 0.4$ 的条件下，多数函数并非最优。这个发现虽然只是在一个非常具体的参数点上，但足以推翻一个长达十多年的普遍猜想。

2.2 一个相关的（已解决的）猜想：多数最不稳定

【原文翻译】

3 一个相关的（已解决的）猜想：多数最不稳定 “多数最不稳定”猜想指出：对于所有线性阈值函数 $f: \{-1,1\}^n \to \{-1,1\}$，所有奇数 n，以及所有 $p \in$，噪声稳定性 $\mathrm{Stab}_\rho[f] := \sum_{S\subseteq[n]} \rho^{|S|} \hat{f}(S)^2$ 由多数函数 Maj_n 最小化。等价地，当 z 从擦除模型中以速率 p 抽取时，有 $\mathrm{Stab}_{p(f)} = E[f(z)^2]$，所以该猜想断言多数函数在LTFs中最小化 $f(z)$ 的 L_2 范数。相比之下，带擦除的NICD问题询问的是多数函数是否在无偏 f 中最大化 L_1 范数 $\Phi_p(f) = E|f(z)|$。虽然 $E|f(z)| \le \sqrt{E[f(z)^2]}$，但据我们所知，这两个优化问题之间没有单向的蕴含关系。在识

别出本文提出的反例后，我们注意到它在精神上与中的反例密切相关，例如与 $\mathrm{sgn}(2x_1+2x_2+x_3+x_4+x_5)$ 相关；两者都具有“两重三轻”的LTF结构，并且在变量的排列和翻转下，它们是同一个函数。

【专业解读】

这一节将新发现置于更广阔的学术背景中，展示了科学思想的传承与关联。作者提到了一个“姊妹问题”，即“多数最不稳定”（Majority is Least Stable）猜想 。这个猜想关注的是另一个衡量函数稳定性的指标，即 L_2 范数 $E[f(z)^2]$，它衡量的是信号的“平均能量”或“功率”。该猜想认为，在所有线性阈值函数中，多数函数的信号能量是最低的（最不稳定）。

有趣的是，那个“最不稳定”猜想也曾被一个反例推翻，而那个反例的结构——例如 $\mathrm{sgn}(2x_1+2x_2+x_3+x_4+x_5)$——与AI这次找到的反例惊人地相似，都是“两重三轻”的加权模式 。这揭示了一个深刻的现象：AI可能并不是凭空“创造”了这个反例，而是通过其强大的模式识别能力，在庞大的函数空间中“重新发现”了一个在相关问题上被证明行之有效的结构。这表明，AI在数学发现中的一个潜在强大作用，是作为一种“模式迁移”的引擎，将在一个领域被证明有效的结构性思想，应用到新的、看似无关的领域中去。人类科学家的任务，则是理解这种迁移为何有效，并从理论上阐明其背后的深层联系。

【原文翻译】

备注 2。由于缺乏正交性，L_1 范数之间的比较通常比 L_2 范数之间的比较更为精细。这在附录A中有所体现：L_1 的手动验证比中相应的 L_2 验证涉及更多的分类讨论。这可能有助于解释为什么一个在“多数最不稳定”背景下熟悉的反例，直到现在才被用于测试带擦除的NICD问题。

备注 3。在“多数最不稳定”问题（L_2 范数）中，反例在 $p=0$ 的一个邻域内有效。另一方面，在带擦除的NICD问题中，对于每个固定的 n，在 $p=0$ 的邻域内没有反例。实际上，在这种情况下，我们证明了多数函数是最优的（见第4节）。

备注 4。对于这个有限问题，暴力搜索是可行的；事实上，GPT-5 Pro在聊天记录中指出它使用了这种方法。尽管如此，能够针对一个开放问题列表进行编程并发现一个具体的反例，在我们看来是值得注意的。

【专业解读】

这几条备注进一步阐明了问题的复杂性和AI贡献的独特性。备注2指出了L_1范数（平均信心）和L_2范数（平均能量）在数学处理上的技术差异，解释了为什么一个相似的结构在之前没有被人们尝试应用到这个问题上。

备注3则揭示了一个关键的区别：在L_2问题中，反例在噪声极小（p 接近0）时就成立了；而在这个L_1问题中，我们稍后会看到，当噪声极小时，多数函数又重新成为了最优选择。这暗示了我们

正在研究的系统存在一种“相变”行为：最优策略会随着环境（噪声水平 p）的变化而改变。

备注4坦诚地指出了AI所使用的方法——“暴力搜索”。对于 $n = 5$ 这样的小维度，理论上可以遍历所有可能的函数结构。但这丝毫没有减弱AI贡献的价值。科学发现的价值不在于方法的“优雅”，而在于结果的“新颖”和“重要性”。在一个被领域内专家研究了十多年的问题上，即使是通过“暴力搜索”找到一个被忽略的角落并发现宝藏，也同样是一项了不起的成就。这表明，AI可以作为一种强大的工具，系统性地扫描那些因过于繁琐而被人类研究者跳过的可能性空间。

第3部分：理论边界

3.1 对于固定的n，在p=0的邻域内多数函数是最优的

【原文翻译】

4 对于每个固定的n，在p=0的邻域内多数函数是最优的

引理 5。对于无偏函数 $f:\{\pm1\}^n \to \{\pm1\}$，我们有 $\Phi_p(f) = p\sum_{i=1}^n |\hat{f}(i)| + O(p^2)$ *证明*。令 $K := |\{i : z_i \neq 0\}|$，所以 $K \sim \mathrm{Bin}(n, p)$。通过对 K 取条件，$\Phi_p(f) = \sum_{k=0}^n \Pr[K = k]E[|f(Z)| \mid K = k]$。我们处理三种情况。情况 $K = 0$。那么 $z = 0$，且 $|f(Z)| = |\hat{f}(\emptyset)| = 0$，因为 f 是无偏的。情况 $K = 1$。我们有 $E[|f(Z)| \mid K = 1] = \frac{1}{n}\sum_{i=1}^n |\hat{f}(\{i\})|$，因此来自 $K = 1$ 的无条件贡献等于 $\Pr[K = 1]\cdot E[|f(Z)| \mid K = 1] = np(1-p)^{n-1}\frac{1}{n}\sum_{i=1}^n |\hat{f}(i)|$。因此 $p\sum_{i=1}^n |\hat{f}(i)| - p(1-(1-p)^{n-1})\sum_{i=1}^n |\hat{f}(i)| = p\sum_{i=1}^n |\hat{f}(i)| + O(p^2)$ 使用了 $1-(1-p)^{n-1} \le (n-1)p$ 和 $\sum_i |\hat{f}(i)| \le \sqrt{n}$ (通过柯西-施瓦茨不等式)。情况 $K \ge 2$。由于 $|f(Z)| \le 1$，$\sum_{k\ge2}\Pr[K = k]E[|f(Z)| \mid K = k] \le \Pr[K \ge 2] \le \binom{n}{2}p^2 + O(p^3) = O(p^2)$。结合这三种情况，得到 $\Phi_p(f) = p\sum_{i=1}^n |\hat{f}(i)| + O(p^2)$，如所声称。

【专业解读】

这一部分从数学上精确地回答了一个问题：反例虽然存在，但它有自己的“适用范围”。这个引理的证明过程非常精彩，它通过对“幸存”比特数量 K 进行分类讨论，揭示了当噪声水平 p 非常非常低时，决定函数表现的主要因素是什么。

- 当 $K = 0$ 时，所有信息都丢失了，输出只能是0，对总体表现没有贡献。
- 当 $K = 1$ 时，只有一个比特幸存。这种情况发生的概率大约与 p 成正比。此时函数的表现完全依赖于它如何处理这单个的信息。
- 当 $K \ge 2$ 时，至少有两个比特幸存。这种情况发生的概率大约与 p^2 成正比。

当 p 是一个非常小的数时（比如0.001），p^2（0.000001）要比 p 小得多。因此，在低噪声环境下，起决定性作用的是 $K = 1$ 的情况。引理的结论 $\Phi_p(f) = p\sum_{i=1}^n |\hat{f}(i)| + O(p^2)$ 精确地

表达了这一点：函数的整体表现 $\Phi_p(f)$ 主要由一个与 p 成正比的项决定，这个项的系数是 $\sum_{i=1}^{n} |\hat{f}(i)|$，即所有单个输入影响力的总和。所有更复杂的情况（$K \geq 2$）都被归入了高阶小量 $O(p^2)$ 中，可以暂时忽略。

【原文翻译】

引理 6 (, Thm. 2.33)。设 n 为奇数。在所有布尔函数 $f: \{-1,1\}^n \to \{-1,1\}$ 中，量 $\sum_{i=1}^{n} |\hat{f}(i)|$ 由多数函数 $\mathrm{Maj}_n(x) = \mathrm{sgn}(x_1 + \cdots + x_n)$ 唯一最大化。

定理 7。固定奇数 n。存在 $p_0 = p_0(n) > 0$ 使得对于所有 $p \in (0, p_0)$，函数 Maj_n 是 Φ_p 在无偏布尔 f 上的唯一最大化者；即，$\Phi_p(\mathrm{Maj}_n) > \Phi_p(f)$ 对于所有无偏 $f \neq \mathrm{Maj}_n$。*证明*。根据引理5，$\Phi_p(g) = p\sum_{i=1}^{n} |g(i)| + E_g(p)$，$|E_g(p)| \leq C_n p^2$，对于每个无偏 g，其中 C_n 是一个常数。由于 Φ_p 和 $\sum_i |g(i)|$ 在翻转单个坐标下是不变的，我们可以不失一般性地假设对于竞争者 f，所有1阶系数都是非负的，所以 $\sum_i |\hat{f}(i)| = \sum_i \hat{f}(i)$，对于 Maj_n 也是如此。令 $\delta_n := \min_{f \neq \mathrm{Maj}_n} \left(\sum_{i=1}^{n} \widehat{\mathrm{Maj}}_n(i) - \sum_{i=1}^{n} \hat{f}(i)\right) > 0$ 根据引理6和在固定 n 下定义域 $\{-1,1\}^n$ 的有限性，这个值是严格为正的（最大化者的唯一性意味着存在一个正的差距）。那么对于任何无偏 $f \neq \mathrm{Maj}_n$，$\Phi_p(\mathrm{Maj}_n) - \Phi_p(f) = p\left(\sum_i \widehat{\mathrm{Maj}}_n(i) - \sum_i \hat{f}(i)\right) + E_{\mathrm{Majn}}(p) - E_f(p) \geq p\,\delta_n - 2C_n p^2$。选择 p_0 足够小。那么对于所有 $p \in (0, p_0)$，我们有 $p\,\delta_n - 2C_n p^2 \geq \frac{1}{2} p\,\delta_n > 0$，证明了该论断和唯一性。

【专业解读】

引理6引用了一个已知的经典结果：在所有布尔函数中，多数函数拥有最大的“总个体影响力”（即 $\sum |\hat{f}(i)|$ 最大）。这符合我们的直觉，多数函数公平地对待每个输入，使得每个输入都有一定的、非零的影响力，其总和是最大的。

定理7则将前面两个引理的结论完美地结合在一起，给出了一个决定性的论断。证明的逻辑链条如下：

1. 从引理5我们知道，当噪声 p 很小时，函数的表现 $\Phi_p(f)$ 主要取决于它的“总个体影响力” $\sum |\hat{f}(i)|$。
2. 从引理6我们知道，在所有函数中，“多数函数”的“总个体影响力”是最大的。
3. 因此，当噪声 p 足够小时，表现最好的函数必然是“多数函数”。

这个证明不仅确认了多数函数在低噪声领域的“王者地位”，还揭示了前面提到的“相变”现象。当 p 很小时，线性项 $p\delta_n$ 在差值比较中占主导，保证了多数函数的胜利。但随着 p 增大，二次项 $-2C_n p^2$ 的负面影响会越来越大，最终可能会超过线性项的优势，从而为其他函数（比如那个反例函数）的胜出创造了机会。这篇论文不仅找到了一个反例，还从理论上阐明了为什么这个反例只在特定的“高噪声”区域出现，而在“低噪声”区域，旧的王者依然无法被撼动。这展现了数学研究的深度和严谨性。

第4部分：细节 - 手动验证

4.1 附录A - 手动验证

【原文翻译】

附录 A 手动验证。 多数函数的的多线性展开。多数函数 $\mathrm{Maj}_n : \{-1,1\}^n \to \{-1,1\}$ 对所有奇数 n 定义为 $\mathrm{Maj}_n(x_1, ..., x_n) = \mathrm{sgn}(x_1 + \cdots + x_n)$，即，如果多数坐标为1，则 $\mathrm{Maj}_n(x) = 1$，否则为-1。其对于 $n = 3, 5$ 的唯一多线性展开为 $\mathrm{Maj}_3(x_1, x_2, x_3) = \frac{1}{2}(x_1 + x_2 + x_3) - \frac{1}{2}x_1x_2x_3, \mathrm{Maj}_5(x) = \frac{3}{8}\sum_{i=1}^{5} x_i - \frac{1}{8}\sum_{1 \le i<j<k \le 5} x_ix_jx_k + \frac{1}{8}x_1x_2x_3x_4x_5.$

反例函数f的多线性展开。由GPT-5 Pro找到的反例，$f(x_1, ..., x_5) = \mathrm{sgn}(x_1 - 3x_2 + x_3 - x_4 + 3x_5)$，具有以下因式分解的多线性形式： $f(x) = \frac{1}{4}(2(-x_2 + x_5) + (1 + x_2x_5)(x_1 + x_3 - x_4 + x_1x_3x_4))$。 事实上，如果 $-x_2 + x_5 = 0$，则输出是 $(x_1, x_3, -x_4)$ 的多数；否则输出是 $(-x_2, x_5)$ 的多数。因此 $f(x) = \mathrm{Maj}_2(-x_2, x_5)$ if $x_2 \neq x_5$，and $\mathrm{Maj}_3(x_1, x_3, -x_4)$ if $x_2 = x_5$。 代入 $\mathrm{Maj}_3(a, b, c) = \frac{1}{2}(a + b + c - abc)$ 即可得到显示的公式。

【专业解读】

这部分是论文的“附录”，是数学家“展示工作”的地方，旨在证明他们的结论不是计算机模拟的侥幸，而是可以被人类逻辑一步步验证的坚实结果。这里给出了计算所需的两个关键“原料”：多数函数 Maj_5 和反例函数 f 的多线性展开式。

Maj_5 的展开式看起来复杂，但它揭示了多数函数的内在结构：它不仅依赖于单个输入（x_i项），还依赖于三个输入的组合（$x_ix_jx_k$项）和所有五个输入的组合。

反例函数 f 的展开式则通过一个巧妙的因式分解，揭示了它的“决策逻辑”。这个逻辑可以这样理解：

1. 首先，检查两个“超级投票者” x_2 和 x_5 的意见是否一致。
2. 如果他们的意见不一致（一个+1，一个-1），那么最终结果就由他们俩的“加权多数”决定（实际上就是由权重更大的 x_5 或 $-x_2$ 决定）。
3. 如果他们的意见一致（都是+1或都是-1），那么他们俩就“退居二线”，让另外三个“普通投票者” $x_1, x_3, -x_4$ 来进行一次标准的3人多数投票，以决定最终结果。 这种分情况讨论的决策机制，正是这个函数能够在特定噪声下表现出色的原因。它是一种动态的、适应性的策略。

【原文翻译】

$|\, Maj_5(Z) \,|$ 的期望值。通过在 $z \in \{-1, 0, 1\}^5$ 处评估其多线性展开，将 $\{-1, 1\}^5$ 上的函数扩展到 $\{-1, 0, 1\}^5$。设 z 从擦除模型中抽取，并设 $K = \#\{i : z_i = 0\} \sim \mathrm{Bin}(5, p)$，

$\Pr(K=k)=\binom{5}{k}p^k(1-p)^{5-k}$。那么 $E|Maj_5(z)| = \sum_{k=0}^{5} E[|Maj_5(z)| \mid K=k]\Pr(K=k) = 0\cdot(1-p)^5 + \frac{3}{8}\cdot 5p(1-p)^4 + \frac{3}{8}\cdot 10p^2(1-p)^3 + \frac{5}{8}\cdot 10p^3(1-p)^2 + \frac{5}{8}\cdot 5p^4(1-p) + 1\cdot p^5 = \frac{15}{8}p - \frac{15}{4}p^2 + \frac{25}{4}p^3 - \frac{45}{8}p^4 + \frac{9}{4}p^5$. 以下是条件期望，通过对称性（我们可以固定哪些坐标非零，并对其符号进行平均）。 情况 $K=0$：$z=(0,0,0,0,0)$，所以 $|Maj_5(z)|=0$。 情况 $K=1$：不失一般性（通过对称性）$z_1 \neq 0$，那么 $|Maj_5(z)| = |\frac{3}{8}z_1| = \frac{3}{8}$。情况 $K=2$：假设 $z_1,z_2 \neq 0$，则 $Maj_5(z)=\frac{3}{8}(z_1+z_2)$，因此条件期望为 $\frac{3}{8}$ 。情况 $K=3$：假设 $z_1,z_2,z_3 \neq 0$，则 $Maj_5(z)=\frac{3}{8}(z_1+z_2+z_3)-\frac{1}{8}z_1z_2z_3$；对 $\{\pm1\}^3$ 进行检查得到 $\frac{5}{8}$。 情况 $K=4$：假设 $z_1,\ldots,z_4 \neq 0, z_5=0$。条件期望为 $\frac{10}{16}\cdot 1=\frac{5}{8}$。 情况 $K=5$ ：显然值为1。

【专业解读】

这是手动验证的第一步：计算基准选手 Maj₅ 的得分。这里的计算策略是“分而治之”，根据输入 z 中有多少个非零元素（即幸存比特的数量 K）来进行分类讨论。下面的表格清晰地整理了这个过程：

表2：Maj₅ 的条件期望计算

K (非零比特数)	样例 z	条件期望 $\|Maj_5(z)\|$	该K值出现的概率	对总期望的贡献
0	(0,0,0,0,0)	0	$(1-p)^5$	0
1	(z_1,0,0,0,0)	3/8	$5p(1-p)^4$	$(15/8)p(1-p)^4$
2	(z_1,z_2,0,0,0)	3/8	$10p^2(1-p)^3$	$(30/8)p^2(1-p)^3$
3	(z_1,z_2,z_3,0,0)	5/8	$10p^3(1-p)^2$	$(50/8)p^3(1-p)^2$
4	(z_1,z_2,z_3,z_4,0)	5/8	$5p^4(1-p)$	$(25/8)p^4(1-p)$
5	(z_1,z_2,z_3,z_4,z_5)	1	p^5	p^5

将表格最后一列的所有项相加，再进行代数化简，就得到了 Maj₅ 表现的最终多项式公式。这个过程虽然繁琐，但每一步都基于基本的概率论和代数运算，是完全可以手动验证的。

【原文翻译】

$|f(z)|$ **的期望值**。设 $M := \frac{1}{2}(z_1+z_3-z_4+z_1z_3z_4)$。使用因式分解形式， $f(z)=\frac{1}{2}(-z_2+z_5)+\frac{1}{2}(1+z_2z_5)M$. 对 (Z_2,Z_5) 取条件期望得到 $E|f(z)| = (1-p)^2E|M| + 2p(1-p)\frac{1}{2}E[|M+1| + |M-1|] + p^2E|M| = (1-2p+2p^2)E|M| + p - \frac{p^2}{2}$，因为四种情况是：两者都为零（概率 $(1-p)^2$）得到 $|M|$；恰好一个为零（概率 $2p(1-p)$），得到 $\frac{1}{4}(|M+1| + |M-1|) = \frac{1}{2}\max(|M|,1) = \frac{1}{2}$ 因为 $M\in[-1,1]$；两者非零且相等（概率 $p^2/2$），得到 $|M|$；两者非零且相反（概率 $p^2/2$），得到 1。 剩下的就是根据 (z_1,z_3,z_4) 中非零元素的数量来计算 $E|M|$： 以概率 $(1-p)^3$ 没有非零元素，则 $|M|=0$；以概率 $3p(1-p)^2$ 恰好有一个，得到 $|M|=\frac{1}{2}$；以概率 $3p^2(1-p)$ 恰好有两个，平均也得到 $\frac{1}{2}$；以概率 p^3 有三个，此时 M 是三比特的多数， $|M|=\frac{5}{8}$。因此， $E|M| = \frac{3}{2}p - \frac{3}{2}p^2 + \frac{5}{8}p^3$ 因此 $E|f(z)| = \frac{7}{4}p - \frac{11}{4}p^2 + 2p^3 - \frac{1}{2}p^4 + \frac{5}{16}p^5$。 在

$p = 0.4$ 时，这给出 $E\left[f(z)\right] = \frac{2689}{6250} = 0.43024$，而之前的计算得出 $E\left[Maj_5(z)\right] = \frac{5363}{12500} = 0.42904$。

【专业解读】

这是手动验证的第二步，也是最关键的一步：计算挑战者 f 的得分。这里的计算技巧更加高明。研究者利用了 f 的特殊结构，先只关注那两个“超级投票者” z_2 和 z_5 的状态，分四种情况讨论：

1. z_2, z_5 都被擦除（都为0）：决策权完全交给另外三个普通投票者。
2. 只有一个幸存：函数的决策变得相对简单。
3. 两个都幸存且意见相同：决策权交给普通投票者。
4. 两个都幸存且意见相左：它们直接决定结果。

通过对这四种情况分别计算，并将它们按概率加权平均，研究者巧妙地将一个复杂的五变量问题，分解成了一系列更简单的子问题。这个过程再次展示了在概率计算中“条件期望”这一工具的威力。最终，他们也得到了一个关于 p 的多项式。

最后一步，就是将 $p = 0.4$ 这个具体的数值分别代入两个函数性能的最终公式中。计算结果显示，$E\left[f(z)\right] = 0.43024$ 确实大于 $E\left[Maj_5(z)\right] = 0.42904$。至此，AI提出的反例得到了完全、严格、可重复的人工验证。整个论证的闭环完成了，从AI的“灵光一闪”，到人类的严谨证明，科学的链条严丝合缝。

第5部分：结论与更广泛的影响

5.1 结论性思考

这份简短而深刻的报告，其核心贡献可以概括为两点：一个“破坏性”的发现和一个“建设性”的证明。

“破坏性”的发现是，通过GPT-5 Pro的辅助，研究者们找到了一个具体的反例 $f(x) = \mathrm{sgn}(x_1 - 3x_2 + x_3 - x_4 + 3x_5)$。这个反例无可辩驳地证明，在带擦除的非交互式相关性蒸馏（NICD）问题中，长期以来被认为是“最优”的多数函数，至少在 $n = 5$ 且擦除概率 $p = 0.4$ 的情况下，并非最优。这推翻了一个持续了十多年的猜想，为该领域的研究开辟了新的方向。

“建设性”的证明是，研究者们并没有止步于找到一个反例。他们进一步从理论上证明，多数函数的“最优性”是有条件的。当信息丢失率（即擦除概率 p）非常低时，多数函数确实是无偏布尔函数中的王者。这揭示了一个更为复杂和有趣的现实：不存在一个“一招鲜，吃遍天”的最优函数。最优策略是随着环境（噪声水平 p）的变化而动态变化的。在低噪声区，民主的“多数

派”策略胜出；在某个中等噪声区，精英式的“加权”策略更优；而在更高噪声区（$p \geq 0.5$），极端的“独裁”策略（只听一个人的）反而成为最佳选择。

从更宏观的视角看，这篇论文的意义远远超出了其在布尔函数分析或信息论领域的具体贡献。它是一个关于科学发现新模式的宣言。它展示了大型语言模型如何从一个仅仅处理文本和代码的工具，演变为一个能够深入特定科学领域、并对悬而未决的难题提出具体、可验证、非凡见解的“研究伙伴”。这项工作激励着未来的科学家们，去思考如何更有效地与AI合作，利用AI强大的搜索和模式识别能力，去探索那些由于组合爆炸或计算复杂性而长期无法触及的科学前沿。它预示着一个人类智慧与机器智能协同进化，共同加速科学发现的新时代的到来。