最近邻表示的信息容量

Kordag Mehmet Kilic*、Jin Sima† 和 Jehoshua Bruck*
*美国加利福尼亚理工学院电气工程系，{kkilic,bruck}@caltech.edu
†美国伊利诺伊大学厄巴纳-香槟分校电气与计算机工程系，jsima@illinois.edu

arXiv:2305.05808v1 [cs.CC] 9 May 2023

摘要——冯·诺依曼计算机架构将计算与存储区分开来。相比之下，大脑采用一种集成架构，其中计算与存储无法区分。受大脑架构的启发，我们提出了一种关联计算模型：存储由 $\mathbb{R}^n$ 中的一组向量定义（我们称之为锚点），计算通过从输入向量收敛至最近邻锚点来执行，输出则是与某个锚点关联的标签。具体而言，本文研究关联计算模型中的布尔函数表示，其中输入为二进制向量，相应输出为最近邻锚点的标签（0 或 1）。该模型中布尔函数的信息容量由两个量决定：(i) 锚点的数量（称为最近邻复杂度（NN）），以及 (ii) 表示锚点各项所需的最大位数（称为分辨率）。我们研究了对称布尔函数，并给出了具有最优最近邻复杂度和分辨率的构造。

一、引言

受大脑结构的启发，我们提出了一种联想计算模型：记忆由 $\mathbb{R}^n$ 中的一组向量定义（我们称之为锚点），计算通过将输入向量收敛到最近邻锚点来完成，输出则是与某个锚点相关联的标签。这一范式与大脑快速分类物体并将其映射到自然语言句法的能力有关。例如，即使是孩子，也能毫不费力地瞬间认出“狗”“猫”或“汽车”。

本文研究联想计算模型中的对称布尔函数表示，其中输入是二进制向量，相应的输出是最近邻锚点的标签（0 或 1）。在该模型中，布尔函数的信息容量与两个量相关：（i）锚点的数量（称为最近邻（NN）复杂度 [5]）；（ii）表示锚点条目的最大比特数（称为分辨率）。形象地说，我们关注表示一个布尔函数所需的“单词”（锚点）数量以及“字母表”（分辨率）的规模。

令 $d(a,b)$ 表示向量 a、$b \in \mathbb{R}^n$ 之间的欧几里得距离。最近邻表示及最近邻复杂度定义如下。

定义 1. 布尔函数 f 的最近邻（NN）表示是由 $\mathbb{R}^n$ 的互不相交子集（P、N）构成的一组锚点，使得对于每个满足 $f(X) = 1$ 的 $X \in \{0,1\}^n$，都存在 $p \in P$，使得对于每个 $n \in N$，$d(X,p) < d(X,n)$，反之亦然。神经网络表示的大小为 $|P \cup N|$。

具体而言，神经网络表示的大小是锚点的数量。自然地，我们关注大小最小的表示。

定义 2。布尔函数 f 的最近邻复杂度，是f的所有最近邻表示中的最小大小，记为 $NN(f)$。

下面给出一些易于验证的双输入示例。我们称 $f(X) = 1$ （或为 0），其对应的锚点则称为正（或负）锚点。

X_1	X_2	与(X)	或(X)	异或(X)
0	0	0	0	0
0	1	0	1	1
1	0	0	1	1
1	1	1	1	0
锚点		0.5 0.5 1 1	0 0.5 0.5	0 0.5 1 1

图1。 2输入布尔与$(X) = x_1 \wedge x_2$，或$(X) = x_1 \vee x_2$，以及异或$(X) = x_1 \oplus x_2$ 函数、它们对应的真值及其神经网络表示。正值和锚点用蓝色表示，负值和锚点用红色表示。

图1中的示例属于一类称为对称布尔函数的布尔函数。令 $|X|$ 表示二进制向量 X 中 1 的数量。

定义 3. 如果布尔函数满足 $f(X) = f(\sigma(X))$，其中 $\sigma(.)$ 可以是任意置换函数，则称其为对称布尔函数。也就是说，对称布尔函数 $f(X)$ 是 $|X|$ 的函数。

我们强调，尽管布尔函数是在二进制向量上求值的，锚点的取值仍可以是实数。不过，这并非必要，因为我们可以将布尔超立方体的所有 2^n 个节点都视为锚点，并根据 f 的取值将它们分配给 P 或 N。这意味着，对于任意 n 输入布尔函数，$NN(f) \leq 2^n$。具体而言，我们定义布尔最近邻复杂度（BNN），即将所有锚点限制为二进制向量时的复杂度。因此，很容易看出，$NN(f) \leq BNN(f) \leq 2^n$ [5]。对于几乎所有布尔函数而言，这个最近邻复杂度上界本质上并不紧。

X_1	X_2	$\lvert X\rvert$	异或(X)	布尔最近邻表示
0	0	0	**0**	$\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ 0 & 1 \\ 1 & 0 \\ \mathbf{1} & \mathbf{1} \end{pmatrix}$
0	1	1	1	
1	0	1	1	
1	1	2	**0**	

图2. 具有最优布尔最近邻表示的双输入异或函数。已知 $BNN(\text{XOR}) = 4$，而 $NN(\text{XOR}) = 3$。

图2以异或函数为例展示了最近邻复杂度与布尔最近邻之间的差异。异或函数推广到 n 个变量后称为奇偶函数。若 X 中 1 的个数为奇数，则奇偶函数 PARITY(X) 的值为 1，即 PARITY(X) $= x_1 \oplus x_2 \oplus \cdots \oplus x_n$ 并且 NN 和布尔最近邻之间存在巨大差距。

定理1（[5]）。设 f 是 n个输入的奇偶函数。那么，

- $NN(f) = n + 1$
- $BNN(f) = 2^n$

神经网络与布尔最近邻之间的根本区别在于用于表示锚点条目的位数。为对此进行量化，我们定义神经网络表示的分辨率。粗略地说，表示一个数所需的位数就是它的分辨率，而一种表示的分辨率则是所有锚点中单个条目的最高分辨率。我们称 $A \in \mathbb{R}^{m\times n}$ 为锚点矩阵，如果每一行对应于一个锚点 a_i，用于表示具有 m 个锚点的n输入布尔函数的神经网络表示。我们指出，不失一般性，可以假设 $A \in \mathbb{Q}^{m\times n}$。

定义 4. 有理数a/b 的分辨率（RES）为 $RES(a/b) = \lceil \max\{\log_2 |a+1|, \log_2 |b+1|\} \rceil$，其中$a$、$b \in \mathbb{Z}$、$b \neq 0$，且它们互质。

对于一个矩阵 $A \in \mathbb{Q}^{m\times n}$, $RES(A) = \max_{i,\,j} RES(a_{ij})$。最近邻表示的分辨率为 $RES(A)$，其中 A是相应的锚点矩阵。

根据定理1中的奇偶函数示例，我们观察到最近邻复杂度与相应分辨率之间存在权衡。 BNN(奇偶函数)在n上呈指数级，分辨率为1；而 NN(奇偶函数)在 n上呈线性级，分辨率为 $\lceil \log_2(n+1) \rceil$。

我们的论文受到 Hajnal、Liu 和 Turán [5] 的工作的启发。他们的论文给出了一些最近邻复杂度结果，包括针对 PARITY 的构造（见定理 1）；然而，对于一般的对称函数，他们只提出了次优的基于 PARITY 的构造（见命题 1）。我们简要提及一些其他相关背景：最近邻分类是信息论和机器学习中一个研究充分的课题 [3, 7]。使用最近邻优化集合表示，最初是在最小化训练集规模的背景下讨论的，即寻找一个最小训练集，使最近邻能够表示原始训练集 [17, 23]。通过最近邻范式表示布尔函数的想法最初在 [4] 中得到研究，当时研究者考虑了通过选择距离度量来优化神经网络表示复杂度的布尔最近邻。这项工作后来扩展到了基于包含的相似度 [18]，并且证明了这种方法能够为析取范式和合取范式公式提供多项式规模的表示。

本文聚焦于对称布尔函数的最近邻表示研究。具体而言，我们为对称布尔函数构造最优规模的最近邻表示。这些函数有助于证明许多关于布尔函数复杂度的结果 [2, 9, 22]；更重要的是，在最近邻的语境下，任何 n输入布尔函数都可以解释为一个 2^n输入的对称布尔函数 [20]。因此，关于对称布尔函数的结果能够帮助我们加深对布尔函数复杂度的理解。此外，我们还研究最近邻复杂度（最近邻表示中锚点的数量）与最近邻分辨率（表示锚点各项所需的最大比特数）之间的权衡。一个违反直觉的发现是，对于对称布尔函数，复杂度和分辨率上的最优性是由各项不对称的锚点实现的！

A. 对称函数的最近邻复杂度

定理1中提到的奇偶函数构造可以扩展到任意对称布尔函数。其思路是为 $|X|$的每个取值分配一个锚点。对于给定的对称布尔函数 $f(X)$，我们可以构造如下带有锚点矩阵 A 的神经网络表示。

$\lvert X\rvert$	$f(X)$	
0	**0**	$\to a_1$
1	**0**	$\to a_2$
2	1	$\to a_3$
3	1	$\to a_4$
4	1	$\to a_5$
5	**0**	$\to a_6$

$$A = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0.2} & \mathbf{0.2} & \mathbf{0.2} & \mathbf{0.2} & \mathbf{0.2} \\ 0.4 & 0.4 & 0.4 & 0.4 & 0.4 \\ 0.6 & 0.6 & 0.6 & 0.6 & 0.6 \\ 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \end{bmatrix} \tag{1}$$

命题 1（[5]）。设 f 是一个具有 n 个输入的对称布尔函数。那么，$NN(f) \leq n + 1$。

命题 1 可通过构造证明。一般而言，矩阵 $A \in \mathbb{R}^{(n+1)\times n}$ 包含锚点 $a_i = ((i-1)/n, \ldots, (i-1)/n)$，其中 P，且 N 是 $\{a_1, \ldots, a_{n+1}\}$ 的一个划分。无论函数本身如何，这种构造都需要 $n+1$ 个锚点，我们称之为基于奇偶函数的构造。相比之下，我们知道有些对称布尔函数所需的锚点数量更少，因此命题 1 显然不是最优的。

可以注意到，对于 AND 和 OR，当按照$|X|$枚举时，$f(X)$ 中都只有一次从 0 到 1 的转变。此外，$f(X)$ 与 $|X|$ 之间的对应关系对于在电路复杂性理论 [12, 13, 19]中衡量对称布尔函数的复杂度尤其有用。这些观察促使我们定义对称布尔函数的区间概念，以衡量最近邻复杂度。

$\lvert X\rvert$	与(X)	或(X)
0	**0**	**0**
1	**0**	1
2	1	1
锚点	$\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0.5} & \mathbf{0.5} \\ 1 & 1 \end{pmatrix}$	$\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ 0.5 & 0.5 \\ 1 & 1 \end{pmatrix}$

图3. 二输入与和或函数的 PARITY 构造。显然，图1给出的构造规模更小，且分辨率相同。

定义5：设 $a \le b \le n$ 为某些非负整数。对于一个 n输入的对称布尔函数，其区间 $[a,b]$ 定义如下：

1) $f(X)$ 对于 $|X| \in [a,b]$ 是常数。2) 如果 $a \neq 0$ 与 $b \neq n$，则对于任意 $|X_1| \in [a,b]$，$f(X_1) \neq f(X_2)$。

- 如果 $a > 0$ 与 $b < n$，$|X_2| = a-1$ 与 $|X_2| = b+1$。
- 如果 $a = 0$ 与 $b < n$，$|X_2| = b+1$。
- 如果 $a > 0$ 且 $b = n$，$|X_2| = a-1$。

量 $I(f)$ 是对称布尔函数 f 的区间总数。

对于一个 n 输入的对称布尔函数，我们通常将其 $[a,b]$ 区间简称为区间。式 1 中的示例函数和 XOR 有 $I(f) = 3$（见图2），而 AND 和 OR 有 $I(f) = 2$（见图3）。

通过为每个区间分配一个锚点来简化命题 1 中的构造，从而将表示的大小缩减至 $I(f)$，这看起来很自然。我们将这称为 $I(f)$ 区间锚点赋值，其中锚点与 $f(X)$ 的区间之间存在一一映射。我们将 i^{th} interval$[a,b]$ 分配给锚点 a_i，并使用 $|X|$的值，分别以 $I_{i-1}+1$ 和 I_i 枚举分配给它的区间的起点和终点。为保持一致性，我们取 $I_0 = -1$。下面的函数是一个示例，其中 $f(X)$ 是一个 6 输入对称函数，且 $I(f) = 3$ 和 (a_1, a_2, a_3) 是一个 $I(f)$ 区间锚点赋值。

$\lvert X\rvert$	$f(X)$		
0	**0**	$\to a_1$	$I_0 + 1 = 0$
1	**0**	$\to a_1$	$I_1 = 1$
2	1	$\to a_2$	$I_1 + 1 = 2$
3	1	$\to a_2$	
4	1	$\to a_2$	$I_2 = 4$
5	**0**	$\to a_3$	$I_2 + 1 = I_3 = 5$

(2)

或许可以扩展命题 1 中给出的构造，通过计算并扰动属于各区间的锚点质心，找到一个 $I(f)$ 区间锚点赋值。我们将这一思路称为基于 PARITY 的扩展。基于 PARITY 的扩展中的锚点是对称的，即彼此相等。考虑式 (1) 中给出的构造。可以通过该矩阵得到一个 3 锚点神经网络表示。

$A' \in \mathbb{R}^{3\times 5}$ 使用矩阵 A。

$\lvert X\rvert$	$f(X)$	
0	**0**	$\to a'_1$
1	**0**	
2	1	
3	1	$\to a'_2$
4	1	
5	**0**	$\to a'_3$

$$A' = \begin{bmatrix} \mathbf{0.1} & \mathbf{0.1} & \mathbf{0.1} & \mathbf{0.1} & \mathbf{0.1} \\ 0.6 & 0.6 & 0.6 & 0.6 & 0.6 \\ \mathbf{1.1} & \mathbf{1.1} & \mathbf{1.1} & \mathbf{1.1} & \mathbf{1.1} \end{bmatrix} \tag{3}$$

遗憾的是，基于 PARITY 的扩展无法应用于所有对称布尔函数。我们在附录 B 中给出了一个不存在基于 PARITY 的扩展示例。直观而言，对称布尔函数对每个输入 x_i 赋予相同的重要性。我们表明，通过打破锚点条目中的对称性，并使整个神经网络表示进一步远离布尔超立方体，可以将表示的规模减小到 $I(f)$。

对于对称布尔函数，$NN(f)$ 是否为 $= I(f)$? 要得出这一结论，我们需要对任意对称布尔函数给出相应的下界。不过，这仍是一个未解决的问题；但我们能够对周期对称布尔函数证明这一点。

定义 6. 如果一个对称布尔函数的每个区间长度都相同，则称其为周期函数，该长度记为 T。T 也称为周期。

更准确地说，若 f 是周期为 T的周期对称布尔函数，那么对于 $(k-1)T \le |X| < kT$，$f(X) = 0$（或为 1），当 k 为奇数时；而 $f(X) = 1$（或为 0），当 k为偶数时。奇偶函数是周期为 $T = 1$的周期对称布尔函数。周期对称布尔函数在其他研究中也很有用（参见 [2] 中的完全二次函数，其中 $T = 2$）。

B. 线性阈值函数的最近邻表示

线性阈值函数是将二进制输入加权求和后送入阈值运算所得的函数，即$\mathbb{1}\{\sum_{i=1}^{n} w_i x_i \ge b\}$，其中 $\mathbb{1}\{.\}$ 是一个指示函数，当内部表达式为真时取值为 1，否则为 0。b 称为阈值，且不失一般性，$w_i \in \mathbb{Z}$。对于非恒定线性阈值函数，利用支持向量的最优超平面论证意味着 $NN(f) = 2$，反过来也成立：任何满足$NN(f) = 2$的布尔函数都是线性阈值函数。这些函数之所以有趣，是因为它们是神经网络的构建模块。

我们将一个具有 $I(f)$ 个 $= 2$ 区间的 n 输入对称线性阈值函数定义为对称布尔函数，即 $f(X) = \mathbb{1}\{|X| \ge b\}$，也就是说，$w_i = 1$ 对所有 i 均成立。有趣的是，根据 b 的取值，这些函数的 BNN 复杂度可能具有不同的阶。

定理 2（[5]）。设 f 是一个阈值为 b的 n输入对称线性阈值函数。那么，

- 如果 $b = n/2$且 n为奇数，则 $BNN(f) = 2$。
- 如果 $b = n/2$且 n为偶数，则 $BNN(f) \le n/2 + 2$。
- 若 $b = \lfloor n/3 \rfloor$, $BNN(f) = 2^{\Omega(n)}$。

如果所有锚点都位于布尔超立方体上，那么最近邻复杂度最高可达 $2^{\Omega(n)}$。不过，在本文中，我们证明了线性阈值函数的双锚点最近邻表示的分辨率受权重分辨率的限制。这意味着，所有对称线性阈值函数都存在常数分辨率的最近邻表示。一般而言，对于线性阈值函数，其分辨率为$O(n\log\ n)$ [1, 8, 14]。

C. 贡献与组织结构

我们的主要贡献如下：

- 针对任意 n 输入的对称布尔函数，我们提出了新的构造，其锚点数量为 $I(f)$，分辨率为$O(\log n)$。
- 对于所有周期性对称布尔函数，这些新构造都具有最优的最近邻复杂度。
- 对于任意 n 输入的对称线性阈值函数，总存在一种在规模和分辨率上均最优的最近邻表示。
- 某些对称布尔函数的最近邻表示要求分辨率达到 $\Omega(\log n)$，因此这些新构造在分辨率上是最优的。

本文的组织结构如下。在第 II 节中，我们给出一种显式构造，从而为任意对称布尔函数的最近邻复杂度给出一个上界。在第 III 节中，我们证明了周期对称布尔函数的一个下界，并得出结论：对于这类函数，$NN(f) = I(f)$。随后，在第 IV 节中，我们研究所构造的神经网络的分辨率，证明其为 $O(\log n)$，并且对于某些对称布尔函数而言是最优的。此外，我们还给出一种在规模和分辨率上均为最优的对称线性阈值函数神经网络表示。附录中包含证明的详细内容。

II. 对称布尔函数最近邻复杂度的上界

本节中，我们给出一种显式构造，用于表示具有 $I(f)$ 个区间的对称布尔函数。

定理 3. 对于一个 n 输入的对称布尔函数 f，$NN(f) \leq I(f)$。

此外，只要 $\epsilon > 0$ 足够小，构造时总可以使用矩阵 $B = \mathbb{1} + \epsilon M$，其中 $B \in \mathbb{R}^{(I(f)-1)\times n}$，且 $\mathbb{1}$ 是全一矩阵，$M \in \mathbb{R}^{(I(f)-1)\times n}$ 是一个满行秩矩阵。

对于任意对称布尔函数，我们首先推导出一个 $I(f)$ 区间锚点赋值的充要条件。

引理 1. 设 f 是一个 n 输入的对称布尔函数，$A \in \mathbb{R}^{I(f)\times n}$ 是一个用于 $I(f)$ 区间的锚点矩阵 - 锚点分配。然后，对于任意 $i \in \{2, \ldots, I(f)\}$ 与 $k < i$，以下条件是必要且充分的：

$$\max_{X\in\{0,1\}^n:|X|=I_k} \sum_{j=1}^{n}(a_{ij} - a_{kj})x_j < \frac{1}{2}\sum_{j=1}^{n}(a_{ij}^2 - a_{kj}^2) < \min_{X\in\{0,1\}^n:|X|=I_{i-1}+1} \sum_{j=1}^{n}(a_{ij} - a_{kj})x_j \tag{4}$$

一般而言，与基于奇偶函数的扩展相比，所有 a_{ij}s 都可能各不相同。不过，我们不能任意选择它们。我们将通过以下观察来证明引理 1。

命题 2. 对于任意具有锚点矩阵 $A \in \mathbb{R}^{I(f)\times n}$ 的 $I(f)$ 区间锚点赋值，$a_{ij} > a_{(i-1)j}$ 对于所有 $i \in\{2, \ldots, I(f)\}$ 和 $j \in \{1, \ldots, n\}$。

我们可以简化引理 1 中给出的充要条件。我们证明，只考察形如 $(i-1, i)$ 的 (k, i) 对，与考察所有满足 $k < i$ 的 (k, i) 对相比，仍能提供充要信息。

引理 2. 设 f 是一个 n 输入的对称布尔函数，$A \in \mathbb{R}^{I(f)\times n}$ 是一个 $I(f)$ 区间锚点赋值的锚点矩阵。那么，对于任意 $i \in \{2, \ldots, I(f)\}$，以下条件是充要条件：

$$\max_{X\in\{0,1\}^n:|X|=I_{i-1}} \sum_{j=1}^{n}(a_{ij} - a_{(i-1)j})x_j < \frac{1}{2}\sum_{j=1}^{n}(a_{ij}^2 - a_{(i-1)j}^2) < \min_{X\in\{0,1\}^n:|X|=I_{i-1}+1} \sum_{j=1}^{n}(a_{ij} - a_{(i-1)j})x_j \tag{5}$$

现在，我们通过找出满足引理 2 的显式锚点集合来证明定理 3。在定理 3 中，B 矩阵实际上是针对 $i \in \{2, \ldots, I(f)\}$和 $j \in \{1, \ldots, n\}$由 $b_{ij} = a_{ij} - a_{(i-1)j}$构造的。直观来说，只要 B的各项差异不大，我们就能找到使公式（5）中的左端和右端成立的锚点。因此，当 $\epsilon > 0$足够小时，$B = \mathbb{1} + \epsilon M$ 就是一个有效的选择。

在定理 3 中，我们指出，满行秩 B 矩阵这一性质只是找到一种构造的充分条件。例如，如果对所有 i都有 $b_{ij} = 1/n$ ， j，那么这种构造可以简化为 [5]中给出的基于 PARITY 的构造。

一般而言，基于 PARITY 的方法会得到 $O(\log n)$ 分辨率。我们可以选取 B ，使得任意对称布尔函数也能通过 $I(f)$ 个锚点达到 $O(\log\ n)$ 分辨率。我们在第 IV 节中给出一族具有 $O(\log n)$ 分辨率的 B 示例。附录 B 中已经给出了这样的示例。

III. 对称布尔函数的最近邻复杂度下界

关于对称布尔函数的电路复杂度下界有多种。类似地，使用布尔函数的 $\{1, 2\}$-符号表示，可以证明 PARITY 所需锚点数量的下界 [5, 6]。我们

证明关于周期对称布尔函数的一个更一般的结果。

定理 4. 对于具有$I(f)$个区间的周期对称布尔函数，$NN(f) \geq I(f)$。

为了证明该定理，我们使用周期对称布尔函数的任何神经网络表示都必须满足的一个必要条件。

命题 3. 考虑一个周期为 T、具有锚点矩阵 $A \in \mathbb{R}^{m\times n}$ 的周期对称布尔函数 $f(X)$ 的神经网络表示。假设对于任意 $i \geq 2$，$\sum_{j=1}^{T} a_{1j} \geq \sum_{j=1}^{T} a_{ij}$。那么，与 a_1 最近的二进制向量不可能在前 T 个坐标上全为 0。

我们首先指出，命题 3 中的假设不失一般性：首先观察到，对称布尔函数的神经网络表示在锚点置换下等价。我们在附录 A2 中明确证明这一点。然后，我们找出条目中 T 总和最大的锚点，并重新排列行和列，使得对于任意 $i \geq 2$，命题 $\sum_{j=1}^{T} a_{1j} \geq \sum_{j=1}^{T} a_{ij}$ 成立。

此外，我们证明，若 $I(f) \leq 4$，则任意 n 输入的对称布尔函数的最近邻复杂度下界为 $NN(f) \geq I(f)$。证明基于第 IV 节中构建的工具，详见附录 C。

IV. 最近邻表示的分辨率

众所周知，基于 PARITY 的方法可为对称布尔函数构造出$\lceil \log_2(n+1) \rceil$ 分辨率。我们声称，第 II 节中提出的构造可实现 $O(\log n)$ 分辨率。设矩阵 A 的 Moore-Penrose 伪逆记为 A^+。

定理 5. 设给定有理矩阵 $B \in \mathbb{Q}^{m\times n}$，满足 $B = \mathbb{1}_{m\times n} + \epsilon I_{m,n}$，其中 $\mathbb{1}_{m\times n}$ 是全一矩阵，ϵ 是有理常数，且 $I_{m,n}$ 是单位矩阵中前 m 行构成的子矩阵。那么，$RES(B^+) = O(\log n)$。

推论 5.1。对于每个 n 输入的对称布尔函数，都存在一个锚点矩阵 $A \in \mathbb{Q}^{I(f)\times n}$，使得 $RES(A) = O(\log n)$。

我们参考 [11] 中的定理 3，显式求出 B^+ 的各项，并证明定理 5。

现在我们考虑线性阈值函数的神经网络表示的分辨率。通过超平面论证，我们证明以下结果，该结果可以通过代数方法验证。

定理 6. 设 f 是一个非恒定的 n 输入线性阈值函数，其权重向量为 $w \in \mathbb{Z}^n$，阈值项为 $b \in \mathbb{Z}$。那么，存在一个分辨率为 $O(RES(w))$ 的 2 锚点神经网络表示。一般而言，分辨率为 $O(n\log n)$。

推论 6.1. 设 f 是一个阈值为 b 的 n 输入对称线性阈值函数。那么，存在一个分辨率为 $O(1)$ 的 2 锚点神经网络表示。

因此，对称线性阈值函数有 2 个区间，并且其神经网络表示在规模和分辨率方面都是最优的。容易验证，

$$a_1 = (0, \ldots, 0, -1/2, -1, \ldots, -1) \tag{6}$$
$$a_2 = (2, \ldots, 2, \ \ 3/2, \ \ 1, \ldots, \ \ 1) \tag{7}$$

始终是一个神经网络表示，其中 a_1 中 0 的数量（以及 a_2 中 2 的数量）为 $b-1$。

然而，如果区间数至少为 3，则对于某些对称布尔函数，分辨率的下界为 $\Omega(\log n)$，并且根据推论 5.1，第 II 节中给出的构造具有最优分辨率。

定理 7. 设 f 是一个 n 输入的对称布尔函数，其中 $f(X) = 1$ 对于 $|X| = \lfloor n/2 \rfloor + 1$ 成立，其他情况下为 0。那么，任何 3 锚点神经网络表示的分辨率均为 $\Omega(\log n)$。

对于定理 7 中的函数，存在一个锚点矩阵 A，使得通过基于 PARITY 的扩展，$RES(A) = O(\log n)$。

为证明定理 7，我们首先证明，除两个例外之外，所有具有 3 个区间的对称布尔函数都需要区间锚点赋值。然后，我们利用引理 2 中的充要条件证明分辨率下界。

引理 3. 设 f 是一个具有 3 个区间的 n 输入对称布尔函数。那么，除函数 $f(X) = 1$（或为 0）在 $|X| \in \{0, n\}$ 时成立、其他情况下为 0（或为 1）之外，任何 $I(f)$ 锚点神经网络表示都必须是区间锚点赋值。

引理 3 中函数的一种替代神经网络表示见附录 C。

最后，我们在附录 D 中利用线性阈值函数的一个电路理论下界，给出了一个依赖于分辨率的神经网络表示规模下界。

V. 总结

我们研究了联想计算模型中对称布尔函数的信息容量。信息容量由两个量构成：最近邻复杂度和分辨率。具体而言，它们分别是神经网络表示中的锚点数量，以及表示这些锚点所需的比特数。具体来说，我们证明了最近邻复杂度的一个上界 $NN(f) \leq I(f)$，并证明该界对周期对称布尔函数是紧的。最近邻复杂度的上界 $I(f)$ 基于一种新的神经网络构造，其分辨率为 $O(\log n)$。我们证明，对于某些对称布尔函数，我们的分辨率上界是紧的。令人惊讶的是，要达到最优性，锚点各项的构造必须是不对称的，这与对称布尔函数的性质相反。此外，我们还给出了对称线性阈值函数在规模 ($NN(f) = 2$) 和分辨率（常数）方面的最优构造。

对于任意对称布尔函数 f，我们猜想 $NN(f) = I(f)$。虽然线性阈值函数具有$NN(f) = 2$，但其分辨率上界为 $O(n\log n)$。一个引人关注的开放问题是为所有.....寻找最近邻表示。

具有多项式神经网络复杂度和对数（或常数）分辨率的线性阈值函数。

致谢

本研究部分由 Carver Mead New AdventureFund 资助。

参考文献

[1] N. Alon 和 V. H. V˜u，“反Hadamard矩阵、硬币称重、阈值门与不可分解超图”，组合理论杂志A辑，第79卷，第1期，第133–160页，1997年。[2] J. Bruck，“多项式阈值函数的调和分析”，SIAM 离散数学杂志，第3卷，第2期，第168–177页，1990年。[3] T. Cover 和 P. Hart，“最近邻模式分类”，IEEE信息论汇刊，第13卷，第1期，第21–27页，1967年。[4] C. Globig 和 S. Lange，“布尔函数类的基于案例的可表示性”，载于第12届欧洲人工智能会议论文集，匈牙利布达佩斯，1996年8月11日至16日，W. Wahlster 编，约翰·威立父子公司，奇切斯特，1996年，第117–121页。[5] P. Hajnal、Z. Liu 和 G. Turán，“布尔函数的最近邻表示”，信息与计算，第285卷，第104 879页，2022年。[6] K. A. Hansen 和 V. V. Podolskii，“多项式阈值函数与布尔阈值电路”，信息与计算，第240卷，第56–73页，2015年。[7] P. Hart，“凝缩最近邻规则（通信）”，IEEE信息论汇刊，第14卷，第3期，第515–516页，1968年。[8] J. Håstad，“阈值门的权重大小”，SIAM 离散数学杂志，第7卷，492，1994 3，484第。– 。[9] J. Håstad、S. Jukna 和 P. Pudlák，“三层电路的自顶向下下界”，计算复杂性，第5卷，第2期，第99–112页，1995年。[10] K. M. Kilic 和 J. Bruck，“使用支配函数的神经网络计算”，载于2021年IEEE国际信息论研讨会（ISIT），IEEE，2021年，第1029–1034页。[11] C. D. Meyer Jr，“修正矩阵的广义逆”，SIAM应用数学杂志，第24卷，第3期，第315–323页，1973年。[12] R. C. Minnick，“线性输入逻辑”，IRE电子计算机汇刊，第1期，第6–16页，1961年。[13] S. Muroga，“多数决策原理下的逻辑元件及其电路复杂性”，载于《美国计算机协会通讯》，计算机协会，纽约百老汇1515号，NY 10036，第2卷，1959年，第14–14页。[14] S. Muroga，《阈值逻辑及其应用》。1971年。[15] R. Paturi 和 M. E. Saks，“奇偶校验阈值电路”，载于《计算机科学基础研讨会论文集》 *[1990] 31*届，IEEE，1990年，第397–404页。[16] V. P. Roychowdhury、A. Orlitsky 和 K.-Y. Siu，“通过通信复杂性得到的阈值电路及相关电路下界”，IEEE信息论汇刊，第40卷，第2期，第467–474页，1994年。[17] S. Salzberg、A. L. Delcher、D. Heath 和 S. Kasif，“最近邻学习的最佳情形结果”，IEEE模式分析与机器智能汇刊，第17卷，第6期，第599–608页，1995年。

[18] K. Satoh，《通过单调理论分析布尔函数的基于案例的可表示性》，载于算法学习理论国际会议，Springer，1998 年，第 179–190 页。[19] K.-Y. Siu 和 J. Bruck，《论小权重阈值电路的能力》，SIAM 离散数学杂志，第 4 卷，第 3 期，第 423–435 页，1991 年。[20] K.-Y. Siu、V. P. Roychowdhury 和 T. Kailath，《神经计算中的深度—规模权衡》，IEEE 计算机汇刊，第 40 卷，第 12 期，第 1402–1412 页，1991 年。[21] D. A. Spielman，《计算任意对称函数》，1992 年。[在线]。网址：https://cpsc.yale.edu/sites/default/files/files/tr906.pdf。[22] L. J. Stockmeyer，《某些对称布尔函数的组合复杂度》，数学系统理论，第 10 卷，第 1 期，第 323–336 页，1976 年。[23] G. Wilfong，《最近邻问题》，载于第七届计算几何年度研讨会论文集，1991 年，第 224–233 页。

附录

A. 定理、引理和命题的证明

1）第 II 节的证明:

命题 2 的证明。考虑两个布尔向量 $X=(x_1,\ldots,0,\ldots,x_n)$ 和 $X'=(x_1,\ldots,1,\ldots,x_n)$，它们仅在 t^{th} 位置不同。假设这发生在一个区间的边界处，即 X 和 X' 分别更接近 a_{i-1} 和 a_i。那么，

$$d(a_{i-1},X)^2-d(a_i,X)^2<0 \tag{8}$$

$$d(a_{i-1},X')^2-d(a_i,X')^2>0 \tag{9}$$

两边不等式相减，得到 $d(a_{i-1},X)^2-d(a_{i-1},X')^2-(d(a_i,X)^2-d(a_i,X')^2)<0$。由于$X$ 与 X' 仅在 t^{th} 位置不同，因此得到 $(2a_{(i-1)t}-1)-2(a_{it}-1)<0$，从而，$a_{it}>a_{(i-1)t}$。□

引理 1 的证明。我们首先写出 $I(f)$ 区间锚点赋值最简单的必要且充分条件。考虑任意两个锚点 a_i 和 a_k，并假设 $|X|\in[I_{i-1}+1,I_i]$，使得 X 更接近 a_i。

$$d(a_i,X)^2-d(a_k,X)^2<0 \tag{10}$$

可以写成

$$\sum_{j=1}^{n}(a_{ij}-x_j)^2-\sum_{j=1}^{n}(a_{kj}-x_j)^2<0 \tag{11}$$

对于 $i\in\{2,\ldots,I(f)\}$ 和 $k<i$。这意味着

$$\frac{1}{2}\sum_{j=1}^{n}(a_{ij}^2-a_{kj}^2) < \min_{X\in\{0,1\}^n:|X|\in[I_{i-1}+1,I_i]}\sum_{j=1}^{n}(a_{ij}-a_{kj})x_j \tag{12}$$

根据命题 2，当 $|X|=\sum_{j=1}^{n}x_j=I_{i-1}+1$ 时，右侧项取最小值，因为每个 $a_{ij}-a_{kj}>0$。因此，

$$\frac{1}{2}\sum_{j=1}^{n}(a_{ij}^2-a_{kj}^2) < \min_{X\in\{0,1\}^n:|X|=I_{i-1}+1}\sum_{j=1}^{n}(a_{ij}-a_{kj})x_j \tag{13}$$

我们对锚点 a_i 和 a_k 进行类似的分析，但这次假设 $|X|\in[I_{k-1}+1,I_k]$，使得 X 比a_k更近。我们仍取 $k<i$。然后，

$$\frac{1}{2}\sum_{j=1}^{n}(a_{kj}^2-a_{ij}^2) < \min_{X\in\{0,1\}^n:|X|\in[I_{k-1}+1,I_k]}\sum_{j=1}^{n}(a_{kj}-a_{ij})x_j \tag{14}$$

$$\frac{1}{2}\sum_{j=1}^{n}(a_{ij}^2-a_{kj}^2) > \max_{X\in\{0,1\}^n:|X|\in[I_{k-1}+1,I_k]}\sum_{j=1}^{n}(a_{ij}-a_{kj})x_j = \max_{X\in\{0,1\}^n:|X|=I_k}\sum_{j=1}^{n}(a_{ij}-a_{kj})x_j \tag{15}$$

现在，当 $|X|=I_k$ 时，右端达到最大值，这由命题 2 得出。结合（13）和（15），证明完成。□

引理 2 的证明。显然，若应用 $k=i-1$，这是必要的。我们将证明，可以通过对公式（5）进行望远镜求和，直到某个固定的 $k<i$，从而得到引理 1。

$$\begin{aligned}
&\max_{X\in\{0,1\}^n:|X|=I_{i-1}}\sum_{j=1}^{n}(a_{ij}-a_{(i-1)j})x_j\\
&\quad<\frac{1}{2}\sum_{j=1}^{n}(a_{ij}^2-a_{(i-1)j}^2)\\
&\quad<\min_{X\in\{0,1\}^n:|X|=I_{i-1}+1}\sum_{j=1}^{n}(a_{ij}-a_{(i-1)j})x_j\\
&\quad\vdots\\
&\max_{X\in\{0,1\}^n:|X|=I_k}\sum_{j=1}^{n}(a_{(k+1)j}-a_{kj})x_j\\
&\quad<\frac{1}{2}\sum_{j=1}^{n}(a_{(k+1)j}^2-a_{kj}^2)\\
&\quad<\min_{X\in\{0,1\}^n:|X|=I_k+1}\sum_{j=1}^{n}(a_{(k+1)j}-a_{kj})x_j
\end{aligned}$$

我们将这些不等式合并，并关注右侧。

$$\sum_{l=k+1}^{i}\min_{X\in\{0,1\}^n:|X|=I_{l-1}+1}\sum_{j=1}^{n}(a_{lj}-a_{(l-1)j})x_j \tag{16}$$

我们可以用 $I_{i-1}+1$ 替换所有最小化表达式的约束集，因为

$$\begin{aligned}
&\min_{X\in\{0,1\}^n:|X|=I_{l-1}+1}\sum_{j=1}^{n}(a_{lj}-a_{(l-1)j})x_j\\
&\quad<\min_{X\in\{0,1\}^n:|X|=I_{i-1}+1}\sum_{j=1}^{n}(a_{lj}-a_{(l-1)j})x_j
\end{aligned} \tag{17}$$

根据命题 2 和 $l\le i$。然后，我们再次将目标函数进行望远镜式组合，使得

$$\frac{1}{2}\sum_{j=1}^{n}(a_{ij}^2-a_{kj}^2) < \min_{X\in\{0,1\}^n:|X|=I_{i-1}+1}\sum_{j=1}^{n}(a_{ij}-a_{kj})x_j \tag{18}$$

左端可以用类似的方法处理。然后，我们可以准确地得到引理 1。□

定理 3 的证明。定义矩阵 $B\in\mathbb{R}^{I(f)-1\times n}$，其中 $b_{ij}=a_{(i+1)j}-a_{ij}$。使用 B 矩阵重写引理 2 中的条件，并将 i 转换为 $i+1$，得到

$$\begin{aligned}
&\max_{X\in\{0,1\}^n:|X|=I_i}\sum_{j=1}^{n}b_{ij}x_j\\
&\quad<\frac{1}{2}\sum_{j=1}^{n}\left(-b_{ij}^2+2b_{ij}\sum_{k=1}^{i}b_{kj}+2b_{ij}a_{1j}\right)\\
&\quad<\min_{X\in\{0,1\}^n:|X|=I_i+1}\sum_{j=1}^{n}b_{ij}x_j
\end{aligned} \tag{19}$$

其中，中间项通过恒等式 $a_{1j} + \sum_{k=1}^{i-1} b_{kj} = a_{1j} - b_{ij} + \sum_{k=1}^{i} b_{kj}$ a_{ij} = 计算得出。

$$\frac{1}{2}\sum_{j=1}^{n}(a_{(i+1)j}^2 - a_{ij}^2) = \frac{1}{2}\sum_{j=1}^{n} b_{ij}(b_{ij} + 2a_{ij}) \tag{20}$$

$$= \frac{1}{2}\sum_{j=1}^{n}\left(b_{ij}^2 + 2b_{ij}\Big(a_{1j} - b_{ij} + \sum_{k=1}^{i} b_{kj}\Big)\right) \tag{21}$$

$$= \frac{1}{2}\sum_{j=1}^{n}\left(-b_{ij}^2 + 2b_{ij}\sum_{k=1}^{i} b_{kj} + 2b_{ij}a_{1j}\right) \tag{22}$$

取左端和右端的凸组合，其中 $\lambda_i \in$ 属于 (0, 1)，使其对每个 $i \in \{1, \ldots, I(f) - 1\}$ 都等于中间项。因此，我们要为 $Ba_1 = c$ 求解 $a_1 =$(a11, . . . , a1n)，其中

$$\begin{aligned}(Ba_1)_i = c_i =& \frac{1}{2}\sum_{j=1}^{n} b_{ij}^2 - \sum_{j=1}^{n} b_{ij}\sum_{k=1}^{i} b_{kj} \\ &+ \lambda_i \max_{X\in\{0,1\}^n : |X| = I_i} \sum_{j=1}^{n} b_{ij}x_j \\ &+ (1-\lambda_i)\min_{X\in\{0,1\}^n : |X| = I_i + 1} \sum_{j=1}^{n} b_{ij}x_j \end{aligned} \tag{23}$$

只要这个线性方程组相容，我们就能得到 a_1 的解，并构造 $I(f)$ 区间-锚点赋值。如果 B 是满行秩矩阵，则该方程组必定相容。只要式 (19) 中左端小于右端，任意满行秩 B 矩阵都适用。因此，我们可以断言，对于充分小的 $\epsilon > 0$，满行秩的 M 可用于为任意对称布尔函数构造锚点矩阵。 □

2）第 III 节的证明:

命题 4. 假设锚点矩阵 $A \in \mathbb{R}^{m\times n}$是一个具有 m 个锚点的神经网络表示，用于一个 n输入的对称布尔函数。那么，对 A 的行和列进行任意置换后，所得矩阵仍是该函数的神经网络表示。

命题 4 的证明. 行的置换显然可行，且适用于任何神经网络表示。对于列，我们可以使用对称布尔函数的定义：对于任意置换 $\sigma(.)$，$f(X) = f(\sigma(X))$。布尔向量 X 的最近锚点可通过以下方式找到

$$\arg\min_i d(a_i, X)^2 = \arg\min_i |X| - 2(AX)_i + ||a_i||_2^2 \tag{24}$$

$$= \arg\max_i \Big(2AX - \mathrm{diag}(AA^T)\Big)_i \tag{25}$$

其中，$||.||_2$ 表示欧几里得范数，diag(M)是一个除对角线元素为 M 外其余元素全为零的矩阵。设 $P \in \{0,1\}^{n\times n}$ 为一个置换矩阵。那么，

$$\arg\max_i \Big(2A(PX) - \mathrm{diag}(AA^T)\Big)_i \tag{26}$$

是分配给同一锚点类型（正类或负类）的锚点索引，因为 f 是对称的。注意，式（ 26 ）等价于

$$\begin{aligned}\arg\max_i \Big(2(AP)X - \mathrm{diag}((AP)(AP)^T)\Big)_i \\ = \arg\max_i \Big(2(AP)X - \mathrm{diag}(AA^T)\Big)_i\end{aligned} \tag{27}$$

□

命题 3 的证明. 我们采用与命题 2 类似的思路。为导出矛盾，假设有一个向量 X =(0, . . . , 0, x_{T+1}, . . . , x_n)被分配给 a_1。令X' =(1, . . . , 1, x_{T+1}, . . . , x_n)。那么，

$$d(a_1, X)^2 < d(a_i, X)^2 \quad \forall i \geq 2 \tag{28}$$

$$d(a_1, X')^2 > d(a_i, X')^2 \quad \text{for some } i \geq 2 \tag{29}$$

根据反证假设，我们得到式 (28)。由于对于给定的 $|X|$，函数值出现了长度为 T 的跳变，因此由 $f(X) \neq f(X')$ 可得式 (29)。将两式相减，得到

$$\sum_{j=1}^{T} a_{1j} < \sum_{j=1}^{T} a_{ij} \tag{30}$$

对于某个 $i \geq 2$，这正是所需的矛盾 ion. □

定理 4 的证明。设 f 是一个周期为 T 的周期对称布尔函数，使得 $n = I(f)T$。假设该函数的神经网络表示具有一个锚点矩阵 $A \in \mathbb{R}^{m\times n}$ ，其中

$$\sum_{j=(k-1)T+1}^{kT} a_{kj} \geq \sum_{j=(k-1)T+1}^{kT} a_{ij} \tag{31}$$

对于任意 $k \in \{1, \ldots, m-1\}$ 与 $k < i$。这不会失去一般性，因为我们可以对锚点的 k个条目的最大和进行排序，并根据命题 4 重新排列A 的行和列。

迭代地，对于每个 $k \in \{1, \ldots, m-1\}$，我们看到，如果对于 $i \in \{1, \ldots, kT\}$而言 $X_i = 0$，那么根据命题 3，($a_1, \ldots, a_k$) 不能被分配给 X 。由于对于周期对称布尔函数而言 $n = I(f)T$ ，如果 $m < I(f)$，则将有 X 个向量（其 $f(X)$ 值不同）被分配给同一个锚点，从而导致矛盾。□

3）第 IV 节的证明:

定理 8（[11]）。假设给定矩阵 $A \in \mathbb{R}^{m\times n}$ 和列向量 $c \in \mathbb{R}^m$、$d \in \mathbb{R}^n$。那么，如果 c 属于 A的列空间，且数量 $\beta = 1 + d^T A^+ c \neq 0$，则有

$$\begin{aligned}(A + cd^T)^+ = A^+ &+ \frac{1}{\beta}v^T k^T A^+ \\ &- \frac{(||k||^2 v^T + k\beta)(||v||^2 k^T A^+ + h\beta)}{\beta(||k||^2||v||^2 + \beta^2)}\end{aligned} \tag{32}$$

where $v = d^T(I - A^+A)$, $k = A^+c$, *and* $h = d^T A^+$.

定理 5 的证明。为应用定理 8，我们取 $c = \mathbb{1}_m$和 $d = \mathbb{1}_n$，其中 $\mathbb{1}_k$ 表示整数 k对应的 $\mathbb{R}^k$ 中的全一向量。此外，取 $A = \epsilon I_{m,\ n}$。因此，$A^+ = \frac{1}{\epsilon} I^T_{m,\ n}$。

我们计算 $\mathbb{1}_{m\times n} + \epsilon I_{m,n}$ 的伪逆如下所示的闭式表达式。

$$\left(\mathbb{1}_{m\times n} + \epsilon I_{m,n}\right)^+ = \begin{bmatrix} \frac{1}{\epsilon} I_{m\times m} - \frac{1+n/\epsilon}{m(n-m)+(m+\epsilon)^2}\mathbb{1}_{m\times m} \\ \frac{1}{m(n-m)+(m+\epsilon)^2}\mathbb{1}_{(n-m)\times m} \end{bmatrix} \tag{33}$$

由于 $m \le n$ 和 ϵ 是有理常数，伪逆所有元素的分母都是 $O(n^2)$ 的函数（而分子则是 $O(n)$ 的函数）。因此，B^+ 各元素的分辨率为 $O(\log n)$。 □

推论 5.1 的证明。对于定理 3 中满足 $m = I(f) - 1$ 的矩阵 $B = \mathbb{1}_{m\times n} + \epsilon M$，我们选取 $\epsilon = 1/2$ 和 $M = I_{m,n}$其中 $I_{m,n}$是单位矩阵的一个子矩阵，包含前 m 行。容易验证，这个 B 矩阵（其中 $b_{ij} = a_{ij} - a_{(i-1)j}$ ）满足区间锚点赋值的充要条件（5）。第一个锚点 a_1 可通过 B^+c 获得，其中 c 见公式（23）。

对于所有 $i \in \{1,\dots,m\}$，借助 $\lambda_i = 1/2$， c 的每个元素都以多项式方式依赖于 B 和 n 上的元素。根据定理 5， B^+ 的元素具有至多 $O(\log n)$ 的分辨率，因此我们可以得出， $a_1 = B^+c$ 的元素具有至多 $O(\log\ n)$ 的分辨率（因此， a_2，.....， a_m）。总之，我们总能得到一个分辨率为 $O(\log n)$ 的区间锚点赋值。 □

定理 6 的证明。从几何上看，线性阈值函数的任意 2 锚点 NN 表示构造都可以写成以下形式：对于实数 $c > 0$ 和任意 $x^* \in \mathbb{R}^n$，使得 $w^T x^* = b'$，其中 $b' = b - 0.5$。

$$a_1 = x^* - cw \tag{34}$$
$$a_2 = x^* + cw \tag{35}$$

这也可以通过代数方式看出。我们将 b 扰动为 b'，以考察超平面本身上的点。我们首先断言，存在 x^* 使得 $RES(x^*) \le RES(w)$。

由于 $f(X)$不是常数，因此总存在一对二进制向量 X'和 X''，使得 $w^T X' < b' < w^T X''$，其中$X'_i = X''_i$对除唯一一个 $i = k \in \mathbb{Z}$之外的所有 $i \in \{1,\dots,n\}$ 均成立。然后，我们构造

$$x^*_i = X'_i \text{ for } i \ne k \tag{36}$$
$$x^*_k = \frac{b' - w^T X'}{w_k} \tag{37}$$

显然， x^*_i除 $i = k$ $|b' - w^T X'| < w_k$ 外均为二进制，因此， $RES(x^*_k) = \lceil \log_2 w_k + 1 \rceil$。总之，$RES(x^*) = \lceil \log_2 w_k + 1 \rceil \le RES(w)$。

取 $c = 1$，我们看到 $RES(A) = O(RES(w))$。此外，由于 $w_i = 2^{O(n \log n)}$一般而言对 $i \in \{1,\dots,n\}$，我们得出 $RES(A) = O(n\log\ n)$。 □

定理 7 的证明。对于定理中给出的 n-输入对称布尔函数，根据引理 3，我们总能得到一个区间-锚点赋值。对于区间-锚点赋值，令 $B \in \mathbb{R}^{2\times n}$ $b_{ij} = a_{(i+1)j} - a_{ij} > 0$ 对 $i \in \{1,2\}$ 以及$j \in \{1,\dots,n\}$。接下来，根据引理 2 写出必要且充分的条件。不失一般性，假设 $0 < b_{11} \le b_{12} \le \cdots \le b_{1n}$，因为我们总能对 B的列重新排序。那么，第一个条件将是

$$\max_{X\in\{0,1\}^n:|X|=\lfloor n/2\rfloor} \sum_{j=1}^n b_{1j}x_j < \frac{1}{2}\sum_{j=1}^n b_{1j}(b_{1j} + 2a_{1j}) < \min_{X\in\{0,1\}^n:|X|=\lfloor n/2\rfloor+1} \sum_{j=1}^n b_{1j}x_j \tag{38}$$

基于此条件，我们有两个重要的不等式。

$$b_{1(n-\sqrt{n})} < b_{1(\sqrt{n}+1)} + \frac{1}{\sqrt{n}} b_{11} \tag{39}$$
$$b_{1n} < 2b_{1\sqrt{n}} \le 2b_{1(\sqrt{n}+1)} \tag{40}$$

为证明式（39），我们假设相反，即 $b_{1(n-\sqrt{n})} \ge b_{1(\sqrt{n}+1)} + \frac{1}{\sqrt{n}} b_{11}$。此外，对于$1 \le j \le \sqrt{n}$， $b_{1(n-\sqrt{n}+j)} \ge b_{1(n-\sqrt{n})}$；对于$1 \le j \le \sqrt{n}+1$， $b_{1(\sqrt{n}+1)} \ge b_{1j}$。因此，若我们对 $j \in \{1,\dots,\sqrt{n}\}$求和，便得到

$$\sum_{j=1}^{\sqrt{n}} b_{1(n-\sqrt{n}+j)} \ge \sum_{j=1}^{\sqrt{n}} \left(b_{1(j+1)} + \frac{1}{\sqrt{n}} b_{11} \right) \tag{41}$$

我们将 $b_{1j}$for $j \in \{n - \lfloor n/2\rfloor + 1, \dots, n - \sqrt{n}\}$ 加到两边。然后，我们得到

$$\sum_{j=n-\lfloor n/2\rfloor+1}^{n} b_{1j} \ge \sum_{j=1}^{\sqrt{n}} \left(b_{1(j+1)} + \frac{1}{\sqrt{n}} b_{11} \right) + \sum_{j=n-\lfloor n/2\rfloor+1}^{n-\sqrt{n}} b_{1j} \tag{42}$$

其中左端有 $\lfloor n/2\rfloor$ 个项，右端有$\lfloor n/2\rfloor + 1$ 个项。这与式（38）矛盾。

类似地，为了证明式（40），我们反设 $b_{1n} \ge 2b_{1\sqrt{n}} \ge b_{11} + b_{12}$。如果我们在两边都加上 $\sum_{i=n-\lfloor n/2\rfloor+1}^{n-1} b_{1i}$，就得到

$$\sum_{i=n-\lfloor n/2\rfloor+1}^{n} b_{1i} \ge b_{11} + b_{12} + \sum_{i=n-\lfloor n/2\rfloor+1}^{n-1} b_{1i} \tag{43}$$

其中，左端有 $\lfloor n/2\rfloor$ 个项，右端有 $\lfloor n/2\rfloor+1$ 个项，这与式 (38) 矛盾。

现在我们希望对式（38）中的中间项进行界定。我们将求和分成三部分。本质上，我们想要证明，整个求和的数值主要由中间项贡献。

$$\sum_{j=1}^{\sqrt{n}} b_{1j}(b_{1j} + 2a_{1j}) + \sum_{j=\sqrt{n}+1}^{n-\sqrt{n}} b_{1j}(b_{1j} + 2a_{1j}) + \sum_{j=n-\sqrt{n}+1}^{n} b_{1j}(b_{1j} + 2a_{1j}) \tag{44}$$

我们还根据 $b_{1j}+2a_{1j}$ 项是否为正，将公式（44）中的中间项分为两部分。令 $\mathcal{J}^+$ 表示满足 $b_{1j}+2a_{1j}>0$ 的索引 j，否则为$\mathcal{J}^-$。然后，我们定义 $\mathcal{S}^+=\sum_{j\in\mathcal{J}^+}($ $b_{1j}+2a_{1j})$ 和 $\mathcal{S}^-=\sum_{j\in\mathcal{J}^-}(b_{1j}+2a_{1j})$，使得

$$\sum_{j=\sqrt{n}+1}^{n-\sqrt{n}} b_{1j}(b_{1j}+2a_{ij}) > b_{1(\sqrt{n}+1)}\mathcal{S}^+ + b_{1(n-\sqrt{n})}\mathcal{S}^- \quad (45)$$

$$> b_{1(\sqrt{n}+1)}\Big(\mathcal{S}^+ + \mathcal{S}^-\Big) - \sqrt{n}2^{r+1}b_{1(\sqrt{n}+1)} \quad (46)$$

其中 r 表示表示的分辨率。如果 r 是表示的分辨率（因此也是锚点矩阵 A的分辨率），显然 $|b_{1j}+2a_{1j}|=|a_{1j}+a_{2j}|\le|a_{1j}|+$ $|a_{2j}|<2^{r+1}$ 39 $b_{1(n-\sqrt{n})}<$，于是我们可以改写公式，使 $b_{1(\sqrt{n}+1)}\Big(1+\frac{1}{\sqrt{n}}\Big)$ 其成为。我们还有一个宽松的界 $\mathcal{S}^->$ $-n2^{r+1}$ 46 ，利用这些结果即可得到公式 () 。

我们通过$-b_{1(\sqrt{n}+1)}2^{r+1}\sqrt{n}$ $-b_{1n}2^{r+1}\sqrt{n}>$ $-2\sqrt{n}b_{1(\sqrt{n}+1)}2^{r+1}$ 求得公式（44）中第一项的下界，并通过 求得第三项的下界。

我们将所有内容合并，所得表达式应满足式 (38) 中的上界。我们再次使用式 (39)。

$$b_{1(\sqrt{n}+1)}\Big(\mathcal{S}^+ + \mathcal{S}^- - \sqrt{n}2^{r+3}\Big)$$

$$< 2\sum_{j=1}^{\lfloor n/2\rfloor+1} b_{1j} < 2\sum_{j=1}^{\lfloor n/2\rfloor+1} b_{1(n-\sqrt{n})}$$

$$< 2b_{1(\sqrt{n}+1)}\Big(\lfloor n/2\rfloor+1\Big)\Big(1+\frac{1}{\sqrt{n}}\Big) \quad (47)$$

$$\mathcal{S}^+ + \mathcal{S}^- < n + O(\sqrt{n}) + \sqrt{n}2^{r+3} \quad (48)$$

将 $\sum_{j=1}^{n}(a_{1j}+a_{2j})=\sum_{j=1}^{n}(b_{1j}+2a_{1j})=$ $\mathcal{S}^+ + \mathcal{S}^- + \sum_{j=1}^{\sqrt{n}} b_{1j}+2a_{1j} + \sum_{j=n-\sqrt{n}+1}^{n} b_{1j}+2a_{1j}$ () () 重写为。

于是，

$$\sum_{j=1}^{n}(a_{1j}+a_{2j}) < \Big(\mathcal{S}^+ + \mathcal{S}^- + 2\sqrt{n}2^{r+1}\Big) \quad (49)$$

根据 $|b_{1j}+2a_{1j}|<2^{r+1}$。令 $r<c\log_2 n$，其中 $0<c<1/2$ 为某个常数。显然，式 (49) 的右侧小于 $n+O(n^\epsilon)$，其中 $1/2<\epsilon<1$ 为一个常数。

类似地，我们得到相应的下界，并使用另一个必要充分条件证明与式 (51) 对应的另一个不等式。对于常数 $1/2<\epsilon<1$，我们有

$$n - O(n^\epsilon) < \sum_{j=1}^{n}(a_{1j}+a_{2j}) < n + O(n^\epsilon) \quad (50)$$

$$n - O(n^\epsilon) < \sum_{j=1}^{n}(a_{3j}+a_{2j}) < n + O(n^\epsilon) \quad (51)$$

两式相减，得到

$$-O(n^\epsilon) < \sum_{j=1}^{n} a_{3j} - a_{1j} < O(n^\epsilon) \quad (52)$$

因此，$0<\frac{1}{2^{2r}}\le|a_{3j}-a_{1j}|<$ $O_{(n^\epsilon)}$ n 对于 $i\in\{1,2\}$和某个 $j\in\{1,\dots,n\}$。因此，$RES(A)=r=\Omega(\log n)$。□

命题 5. 对于任意大小至少为 2 的神经网络表示，任选一个正锚点和一个负锚点，a 和 b。令$\mathcal{X}$ 和 $\mathcal{Y}$ 分别为最接近 a 和 b的二进制向量集合。那么，$conv(\mathcal{X})\cap conv(\mathcal{Y})=\emptyset$ 其中 $conv(\mathcal{A})$ 表示集合 $\mathcal{A}$ 的凸包。

命题 5 的证明。设 $|\mathcal{X}|=k$ 和 $|\mathcal{Y}|=l$。我们假设 $\mathcal{X}=\{X_1,\dots,X_k\}$ 和 $\mathcal{Y}=\{Y_1,\dots,Y_l\}$。此外，设 u 和 v 分别为集合 $\mathcal{X}$ 和 $\mathcal{Y}$ 的任意凸组合。也就是说，$u=\sum_{i=1}^{k}\lambda_i X_i$ 和 $v=\sum_{i=1}^{l}\mu_i Y_i$，其中 $X_i\in\mathcal{X}$、$Y_i\in\mathcal{Y}$ $\lambda_i\in[0,1]$ $\mu_i\in[0,1]$ 、$\sum_{i=1}^{k}\lambda_i=1$，以及 $\sum_{i=1}^{l}\mu_i=1$。

我们知道，对于 $i\in\{1,\dots,k\}$，$d(a,X_i)^2<d(b,X_i)^2$；对于 $i\in\{1,\dots,l\}$，$d(a,Y_i)^2>d(b,Y_i)^2$。反设存在一组 λ 和 μ，使得 $u=v$。

$$d(a,u)^2 - d(b,u)^2 = ||a||^2 - ||b||^2 - 2(a-b)^T\sum_{i=1}^{k}\lambda_i X_i \quad (53)$$

$$= \sum_{i=1}^{k}\lambda_i\Big(||a||^2 - 2(a-b)^T X_i - ||b||^2\Big) \quad (54)$$

$$= \sum_{i=1}^{k}\lambda_i\Big(d(a,X_i)^2 - d(b,X_i)^2\Big) < 0 \quad (55)$$

其中我们使用 $\lambda_i\ge0$ 和 $\sum_{i=1}^{k}\lambda_i=1$。我们对 v 也进行同样的操作，因为 $u=v$，从而得到

$$d(a,v)^2 - d(b,v)^2 = \sum_{i=1}^{l}\mu_i\Big(d(a,Y_i)^2 - d(b,Y_i)^2\Big) > 0 \quad (56)$$

从而导致矛盾。□

引理 3 的证明。对于具有 3 个区间的对称布尔函数，除区间锚点赋值外，仅有两种可能的锚点赋值：要么将 a_2 赋给区域 $|X|\in[I_1+1,I_2]$，而 a_1 与 a_3 共享其余区域（情形 1）；要么 a_2 与 a_3 共享同一区域 $|X|\in[I_1+1,I_2]$，而将 a_1 赋给其余区域（情形 2）。

对于情形 1：我们选取 X_1，并将 X_2 分别赋给 a_1 或 a_3，其中 $|X_1|=t_1<I_1+1$ 和 $|X_2|=t_2>I_2$。另外，独立考虑一个整数 $t\in[I_1+1,I_2]$。除描述中给出的函数外，这始终可行。然而，对于该函数，此情形意味着区间锚点赋值，因为对于 $|X|=0$ 和$|X|=n$ 都存在唯一向量。我们将 $X_1=X_2=$为 1 的坐标集合记为 $\mathcal{S}_1$，将 $X_1=1$、$X_2=0$ 记为 $\mathcal{S}_2$，将 $X_1=0$、$X_2=1$ 记为 $\mathcal{S}_3$，最后将 $X_1=X_2=0$ 记为 $\mathcal{S}_4$。显然，$t_1=|\mathcal{S}_1|+|\mathcal{S}_2|$、$t_2=|\mathcal{S}_1|+|\mathcal{S}_3|$、$|\mathcal{S}_1|+|\mathcal{S}_2|+|\mathcal{S}_3|+|\mathcal{S}_4|=n$。下面给出一个例子。

$$X_1 = (1,\dots,1,1,\dots,1,0,\dots,0,0,\dots,0) \quad (57)$$

$$X_2 = (\underbrace{1,\dots,1}_{\mathcal{S}_1},\underbrace{0,\dots,0}_{\mathcal{S}_2},\underbrace{1,\dots,1}_{\mathcal{S}_3},\underbrace{0,\dots,0}_{\mathcal{S}_4}) \quad (58)$$

注意，通过重新排列索引，可以不失一般性地假设示例表示为公式（57）和（58）所给出的形式。此外，根据 t_1、t_2 和 t 的选择，$\mathcal{S}_3 \neq \emptyset$ 是必要的。显然，$|\mathcal{S}_3| = t_2 - t_1 + |\mathcal{S}_2| \geq t - t_1$ 和 $|\mathcal{S}_3| = t_2 - |\mathcal{S}_1| \geq t - |\mathcal{S}_1|$。

可以轻松验证，凸组合 $X' = \frac{t_2-t}{t_2-t_1}X_1 + \frac{t-t_1}{t_2-t_1}X_2$ 位于超平面 $|X| = t$ 上。我们进一步声称，它属于二进制向量 Xs 的集合的凸包，其中 $|X| = t$。

令 Y 为超平面 $|X| = t$ 上所有二进制向量的平均值，这些向量在 X_1 中 1 的位置有 t_1 个 1，在 X_1 中 0 的位置有 S_4 个 0。也就是说，我们将剩余的 $t - t_1$ 个 1 放在 $\mathcal{S}_3$ 中的索引处。显然，这样的向量共有 $\binom{S_3}{t-t_1}$ 个。$\mathcal{S}_3$ 区域中的平均值为 $K = \binom{|\mathcal{S}_3|-1}{t-t_1-1} / \binom{|\mathcal{S}_3|}{t-t_1} = (t-t_1)/|\mathcal{S}_3|$。

类似地，我们将 Z 定义为超平面 $|X| = t$ 上所有二进制向量的平均值：这些向量在索引 $\mathcal{S}_1$ 处均为 1，而剩余的 $t - |\mathcal{S}_1|$ 个 1 分布在区域 $\mathcal{S}_3$ 中。区域 $\mathcal{S}_3$ 中的平均值为 $L = \binom{|\mathcal{S}_3|-1}{t-|\mathcal{S}_1|-1} / \binom{|\mathcal{S}_3|}{t-|\mathcal{S}_1|} = (t - |\mathcal{S}_1|)/|\mathcal{S}_3|$。

$$Y = (1, \ldots, 1, 1, \ldots, 1, K, \ldots, K, 0 \ldots, 0) \tag{59}$$

$$Z = (\underbrace{1, \ldots, 1}_{\mathcal{S}_1}, \underbrace{0, \ldots, 0}_{\mathcal{S}_2}, \underbrace{L, \ldots, L}_{\mathcal{S}_3}, \underbrace{0, \ldots, 0}_{\mathcal{S}_4}) \tag{60}$$

然后，容易验证 $X' = \frac{t_2-t}{t_2-t_1}Y + \frac{t-t_1}{t_2-t_1}Z$，因为 $K\ (t_2-t)\ /\ (t_2-t_1)\ + L\ (t-t_1)\ /\ (t_2-t_1)\ =\ (t-t_1)\ /\ (t_2-t_1)$。

对于情形 2：当 a_2 和 a_3 共享区间 $[I_1+1, I_2]$ 时，对于 $t \in [I_1+1, I_2]$，所有满足 $|X| = t$ 的向量都不能分配给某个锚点，否则问题将归结为情形 1。因此，对于任意 $t \in [I_1+1, I_2]$，a_2 和 a_3 必须共享 $|X| = t$。

令 $X_1 \neq X_2$，并令 $\mathcal{S}_i$ 与情形 1 中相同，但 $|X_1| = |X_2| = I_1 + 1$除外。我们可以分别选择最接近a_2 或 a_3 的 X_1 和 X_2。在这种情况下，$\mathcal{S}_2$ 和 $\mathcal{S}_3$ 均非空。我们为 X_1&X_2 定义 X_3，使得 $|X_3| = I_2$，并且 X_3 在 $\mathcal{S}_1$ 和 $\mathcal{S}_2$ 中的索引对应的值全为 1，但 $\mathcal{S}_2$ 的第一个索引对应的值为 0。此外，我们也将 $\mathcal{S}_3$ 的第一项设为 0，其余索引的值任意填充。因此，根据构造，我们有 $I_2 < n - 1$。

$$X_3 = (\underbrace{1, \ldots, 1}_{\mathcal{S}_1}, \underbrace{0, 1, \ldots, 1}_{\mathcal{S}_2}, \underbrace{0, 1, 0, \ldots, 1}_{\mathcal{S}_3 \cup \mathcal{S}_4}) \tag{61}$$

类似地，通过对调 0 和 1 的角色，我们可以得到$I_1 > 0$ 条件。因此，只要 $I_1 > 0$ 或 $I_3 < n - 1$成立，就可以构造 X_3。唯一的例外是：当 $f(X)$ 对于 $|X| \in \{0, n\}$ 为 $= 1$（或 0），而对于其他情况为 0（或 1）时。

注意，X_3 是通过使用 $\mathcal{S}$ 构造的，同时取决于 a_2 和 a_3。一般来说，X_3 可以赋值为 a_2或 a_3。假设赋值为 a_2。否则，在以下论断中将X_1 改为 X_2：令 $Y = X_1 \wedge X_3$ 和 $Z = X_1 \vee X_3$。显然，$|Y| \leq I_1 - 1$和 $|Z| \geq I_2 + 1$，且它们都赋值为 a_1。最后，我们断言 $Y/2 + Z/2 = X_1/2 + X_3/2$。

总之，这两种情况都与命题 5 矛盾，因此必须进行区间锚点分配。 □

B. 基于奇偶函数的对称布尔函数神经网络表示扩展的一个反例

设 $f(X)$ 是公式（62）中给出的函数。为了导出矛盾，我们假设存在此对称函数的基于 PARITY 的扩展，即 $a_{ij} = a_{ik}$ 对任意 $i \in \{1, \ldots, 5\}$ 和 $j,\ k \in \{1, \ldots, 8\}$。

$\lvert X\rvert$	$f(X)$	
0	1	$I_1 = 0$
1	**0**	$I_2 = 1$
2	1	
3	1	
4	1	
5	1	
6	1	$I_3 = 6$
7	**0**	$I_4 = 7$
8	1	$I_5 = 8$

(62)

我们可以改写引理 2 中给出的充要条件。我们还知道，对于所有 $j \in \{1, \ldots, 8\}$和 $i \in \{2, \ldots, 5\}$，$a_{ij} + a_{(i-1)j}$（或 $a_{ij} - a_{(i-1)j}$）的值都相同。于是得到

$$\begin{aligned} I_{i-1}(a_{i1} - a_{(i-1)1}) &< 4(a_{i1} - a_{(i-1)1})(a_{i1} + a_{(i-1)1}) \\ &< (I_{i-1} + 1)(a_{i1} - a_{(i-1)1}) \end{aligned} \tag{63}$$

$$\frac{I_{i-1}}{4} < (a_{i1} + a_{(i-1)1}) < \frac{I_{i-1} + 1}{4} \tag{64}$$

回想一下，命题 2 在这里仍然适用。因此，对于所有 $j \in \{1, \ldots, 8\}$，$a_{1j} < a_{2j} < a_{3j} < a_{4j} < a_{5j}$。更明确地说，我们有以下不等式组

$$a_{11} < a_{21} < a_{31} < a_{41} < a_{51} \tag{65}$$

$$0 < a_{21} + a_{11} < \frac{1}{4} \tag{66}$$

$$\frac{1}{4} < a_{31} + a_{21} < \frac{2}{4} \tag{67}$$

$$\frac{6}{4} < a_{41} + a_{31} < \frac{7}{4} \tag{68}$$

$$\frac{7}{4} < a_{51} + a_{41} < \frac{8}{4} \tag{69}$$

首先，我们将公式（66）和（68）分别乘以 -1，并将公式（66）、（67）、（68）和（69）相加。其次，我们将公式（67）乘以 -1，再加到公式（68）上。因此，

$$0 < a_{51} - a_{11} < 1 \tag{70}$$

$$1 < a_{41} - a_{21} < 1.5 \tag{71}$$

这与公式（65）矛盾。

以下是使用我们的方法构造公式（62）中函数的过程。这里我们使用 2 位有效数字来拟合矩阵。

$$\begin{bmatrix} 17.78 & 2.28 & -5.72 & -21.22 & -1.53 & -1.53 & -1.53 & -1.53 \\ 19.28 & 3.28 & -4.72 & -20.22 & -0.53 & -0.53 & -0.53 & -0.53 \\ 20.28 & 4.78 & -3.72 & -19.22 & 0.47 & 0.47 & 0.47 & 0.47 \\ 21.28 & 5.78 & -2.22 & -18.22 & 1.47 & 1.47 & 1.47 & 1.47 \\ 22.28 & 6.78 & -1.22 & -16.72 & 2.47 & 2.47 & 2.47 & 2.47 \end{bmatrix} \tag{72}$$

C. 不含区间锚点分配的对称布尔函数的最近邻表示示例

引理 4. 设 f 是一个 n 输入的对称布尔函数，满足 $I(f) \leq 4$。那么，$NN(f) \geq I(f)$。

证明。当 $I(f) \in \{1,2\}$ 时，结论显然。当 $I(f)$ =3 时，可以使用命题 2。现在，假设对于满足 $I(f) = 4$ 的对称布尔函数，存在 3 锚点神经网络表示。那么，锚点在这些区间上的分配只能属于以下两种情形：a_1 被分配到第一个和第三个区间，而 a_2 与 a_3 共享第二个和第四个区间的二进制向量；或者反之。这两种情形都具有与引理 3 证明中情形 1 类似的结构。按照类似的步骤，我们可以得出与命题 5 矛盾的结论，从而证明对于满足 $I(f) = 4$ 的对称布尔函数，不存在 3 锚点最近邻表示。

□

这里给出一个 5 输入反例，针对引理 3 中描述的函数，其中中间区间由两个正锚点共享。

$\lvert X\rvert$	$f(X)$
0	**0**
1	1
2	1
3	1
4	1
5	**0**

$$A' = \begin{bmatrix} 0 & 0.57 & 0.57 & 0.57 & 0.57 \\ \mathbf{0.5} & \mathbf{0.5} & \mathbf{0.5} & \mathbf{0.5} & \mathbf{0.5} \\ 1 & 0.43 & 0.43 & 0.43 & 0.43 \end{bmatrix} \tag{73}$$

D. 神经网络表示与线性阈值电路之间的关系

神经网络表示的另一个重要方面是它们在电路类层级中的位置。假设给定一个布尔函数的 m-锚点神经网络表示。可以构造一个深度为 3、包含 $O(m^2)$ 个线性阈值门的线性阈值电路 [5, 6]。然而，反过来则尚无定论；给定逻辑电路，我们无法轻易得到神经网络表示。

我们聚焦于比较函数（记作 COMP），以说明规模与分辨率之间的权衡。COMP(X,Y) 是一个线性阈值函数 COMP$(X,Y) = \mathbb{1}\left\{\sum_{i=1}^{n} 2^{i-1}x_i \geq \sum_{i=1}^{n} 2^{i-1}y_i\right\}$，用于判断一个无符号整数是否大于或等于另一个整数。这里，X 和 Y 分别对应这些整数的二进制展开。

令 OR 表示布尔或函数，AND 表示布尔与函数，THR 表示线性阈值函数。

引理 5. 假设 $f(X)$ 是一个 n 输入布尔函数，其神经网络表示包含 m 个锚点，正锚点（或负锚点）集合为 P（或 N）。那么，存在一个逻辑电路，其形式为 OR $\circ$ AND $\circ$ THR，规模为 $|P||N| + |P| + 1$，可以计算 $f(X)$，其中 $\circ$ 表示函数复合。

证明。最近邻的标签可以通过以下公式求得。

$$\arg\min_i d(a_i, X)^2 = \arg\min_i |X| - 2(AX)_i + ||a_i||_2^2 \tag{74}$$

$$= \arg\max_i 2(AX)_i - ||a_i||_2^2 \tag{75}$$

其中 $||.||_2$ 表示欧几里得范数。令 P（或 N）为正（或负）锚点集合 $\{a_1, \ldots, a_{|P|}\}$（或，$\{b_1, \ldots, b_{|N|}\}$。

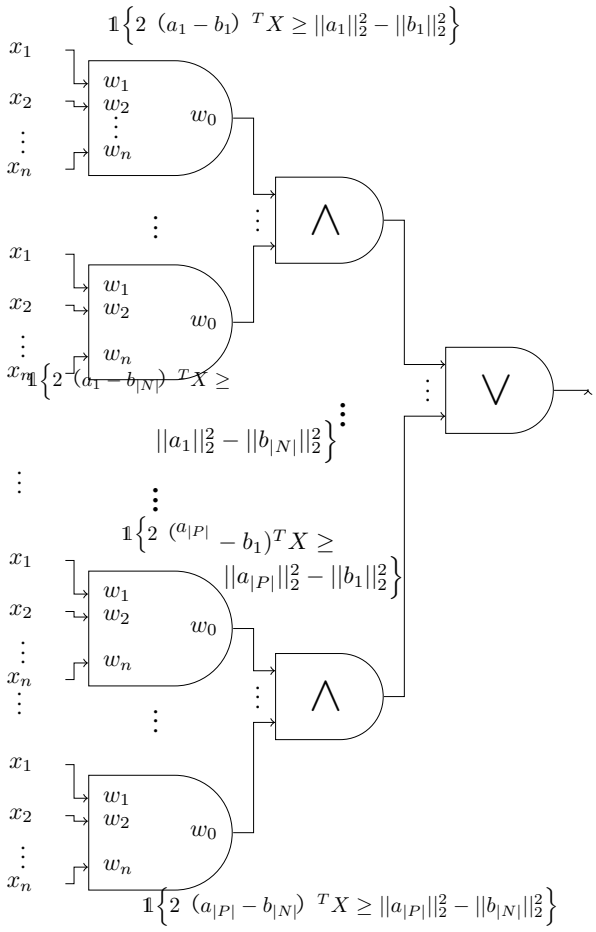

图 4. 将神经网络表示转换为深度 3 线性阈值电路的示意图。第一层由线性阈值门组成，而第二层和第三层构成一个与–或网络。内部带有 w_i、偏置为 w_0 的门是 THR(X)，内部带有 $\wedge$ 的门是与(X)，内部带有 $\vee$ 的门是或(X)

构造性变换如图 4 所示。在第一层中，我们只需使用式 (74) 比较每个正锚点（例如 a_i）到所有负锚点 $\{b_1, \ldots, b_{|N|}\}$ 的距离。因此，第一层中有 $|P||N|$ 个阈值门。每个正锚点对应一个与门，顶层门是一个或门，这表明所声称的电路规模是正确的。

首先假设 $f(X) = 1$。线性阈值函数

$$\mathbb{1}\left\{2a_i^T X - ||a_i||_2^2 \geq 2b_j^T X - ||b_j||_2^2\right\} \tag{76}$$

对于某个 $i \in \{1,\ldots,|P|\}$，以及所有 $j \in \{1,\ldots,|N|\}$，其取值为 1，因为正锚点必须比任何负锚点都更接近 X。因此，相应的 i^{th} 与门的输出将为 1，进而电路的输出也为 1。

反过来，假设 $f(X) = 0$。那么，总存在某些 j，使得对于所有 i，式（76）都为 0，因为距 X 最近的邻居是负的（即位于 N中）。因此，所有第二层与门都输出 0，电路的输出也为 0。 □

在计算 COMP 时，若存在权重大小约束，则阈值门数量存在一个与电路深度无关的下界。根据该下界和引理 5，我们可以得到最近邻复杂度的一个类似下界。

定理9（[10, 16]）。设 $X = (x_1,\ldots,x_n)$ 和 $Y = (y_1,\ldots,y_n)$ 为二进制向量。COMP(X,Y) 函数可以由一个大小为 $\Omega(n/\log nW)$ 的线性阈值电路计算，其中 $W \in \mathbb{Z}$ 为最大权重大小。

引理6。对于一个 n 输入的 COMP(X,Y) 的神经网络表示，其锚点数量为 $\Omega\left(\sqrt{}^{n}{}_{\log n+r}\right)$，其中 ~~$r$~~ 是该表示的分辨率。

证明。假设存在一个 COMP(X,Y) 的神经网络表示，其锚点数量（记为 m）为 $o\left(\sqrt{}^{n}{}_{\log n+r}\right)$。那么，根据引理 5，存在一个大小为 $o\left(\sqrt{}^{n}{}_{\log n+r}\right)$ 的深度为 3 的阈值电路计算 COMP(X,Y)。由于 $r = \lceil \log_2 W + 1 \rceil$，这与定理9中给出的 COMP$(X,Y)$ 的下界矛盾。 □