arXiv:2601.00745v1 [quant-ph] 2 Jan 2026

量子回归的免训练认证界：一种可扩展框架

Demerson N. Gon˛calves[1*]、TharsoD.Fernandes[2]，
Pedro H.G. Lugao[3]，Jo˜ao T.Dias[4]

[1*]数学系，塞尔苏·苏科夫·达·丰塞卡联邦技术教育中心（CEFET-RJ），巴西里约热内卢州彼得罗波利斯。[2]数学系，圣埃斯皮里图联邦大学（UFES），巴西圣埃斯皮里图州阿莱格雷。[3]计算机工程系，CEFET-RJ，巴西里约热内卢州彼得罗波利斯。[4]电信工程系，CEFET-RJ，巴西里约热内卢州里约热内卢。

*通讯作者。电子邮箱：demerson.goncalves@cefet-rj.br；贡献作者：tharso.fernandes@ufes.br；pedro.lugao@cefet-rj.br；joao.dias@cefet-rj.br；

摘要

我们提出了一种无需训练的量子回归认证误差界，该界直接由泡利期望值推导而来。我们将分类中最低准确率的启发式方法推广到回归，并在泡利特征空间内评估轴对齐预测器。我们从形式上证明，最优轴对齐预测器构成一个严格上界，限定了在同一量子特征映射上定义的任何线性或基于核的回归器所能达到的最低训练均方误差（MSE）。由于计算这一精确上界需要对全连接泡利基进行难以处理的扫描，我们引入了一种蒙特卡洛框架，利用一组可处理的测量轴子集对其进行高效估计。此外，我们还提供非渐近统计保证，以便在实际测量预算内认证性能。该方法能够快速比较量子特征映射并及早诊断其表达能力，从而在部署复杂度更高的模型之前，为架构选择提供依据。

关键词：量子核回归、泡利分解、蒙特卡洛采样、特征映射评估。

1 引言

机器学习（ML）能够从高维数据中提取可用于行动的模式，因此已成为不确定性下决策和复杂系统自动化发现的标准范式 [1, 2]。量子机器学习（QML）是一个快速发展的领域，旨在利用量子系统的独特结构和计算资源，突破经典机器学习的局限 [3, 4]。最有前景的 QML 范式之一依赖于量子核方法。该框架的核心是特征映射，这是一种将经典输入数据嵌入高维希尔伯特空间的变换，使复杂的非线性关系能够通过线性分析加以处理。在量子情境下，该映射通过参数化量子电路将数据编码为量子态，从而使经典算法能够在由此产生的量子诱导特征空间中运行。这一量子–经典混合框架由 Havlıˇcek 等人 [5], 、Schuld 和 Killoran [6], 以及 Mitarai 等人 [7]的奠基性研究确立。在这些开创性工作的基础上，后续研究广泛探讨了这些映射的硬件实现和理论性质 [8–11]。

然而，在实践中利用这些量子特征空间面临着诸多重大障碍。在变分方法中，参数化量子电路的优化经常受到贫瘠高原现象的阻碍：随着量子比特数量增加，代价函数的梯度会呈指数级消失，使得深层架构实际上无法训练 [12–14]。为了在不进行完整训练的情况下探索这一领域，人们提出了几何描述符，例如表达能力 和纠缠能力 [15]，用以刻画试探态结构。尽管这些与任务无关的指标能够为状态的统计分布提供有价值的见解，却无法直接预测特定目标函数的泛化性能或训练误差。此外，尽管量子核方法避免了基于梯度的优化，但在含噪中等规模量子（NISQ）时代，它们面临严重的可扩展性限制 [16], ：受限的相干时间和硬件噪声使得估算全连接核矩阵 $N \times N$ 的成本高得令人难以承受。验证量子模型还需要将其与稳健的经典基线进行基准比较，例如标准支持向量回归（SVR） [17], ，这为模型性能设定了很高的门槛。

因此，若要通过对每种候选架构都完整训练模型来评估和比较量子特征映射，计算上将难以实现。在基于核的回归中，这通常需要构造完整的核矩阵并调整正则化超参数；在变分方法中，则需要反复评估量子电路以进行梯度更新。这些开销会随着数据集规模和量子电路深度的增加而迅速恶化。为应对分类中的类似挑战，Suzuki等人 [18]提出了最低准确率启发式方法。该方法通过将测量限制在轴对齐泡利可观测量上，估计可达到的基准性能，从而避免显式优化。然而，原始方法仅针对分类，并依赖于对泡利基的全面扫描。这造成了两个显著的缺口：第一，回归任务没有直接对应的方法，而此类任务的核心是最小化均方误差（MSE）等连续误差；第二，完整泡利分解的计算成本对于 n 个量子比特会按 4^n 的规模增长，因此对于较大的系统而言，精确度量难以计算。

在本研究中，我们从三个主要方面拓展特征映射诊断的分析及其实用性，将轴对齐概念推广到量子回归。首先，我们在泡利特征空间中正式定义回归设定，并提出一种基于单坐标最小二乘的免训练评分。与几何启发式方法不同，该度量直接基于均方误差的最小化。其次，我们给出严格的理论结果（定理 1），证明最佳轴对齐预测器的均方误差构成了同一特征映射下任何线性或基于核的回归器所能达到的最低训练均方误差的认证上界。这从形式上证明了该评分是一个“保证极限”：如果轴对齐误差较低，就能从数学上保证完整量子模型的表现至少同样好。第三，我们提出一个具有统计认证的蒙特卡洛（MC）框架，使该上界在计算上切实可行。我们不再进行难以计算的 4^n 扫描，而是引入一个基于轴的随机子集计算的估计量，$\widehat{\text{均方误差}}_{\text{axis}}$。我们基于Hoeffding集中不等式推导出非渐近概率保证 [19]，，将样本量与上界的可靠性联系起来（定理3）。

从实践角度来看，我们的结果将这一理论上界转化为一种可扩展的工具，用于预筛选量子特征映射。所提出的工作流程无需对完整特征空间进行表征，从而避免了指数级开销。通过随机抽取 $t \ll 4^n$ 个泡利轴并计算简单的一维回归评分，即可获得模型潜在误差的认证上界。这使我们能够快速比较架构、及早诊断表达能力问题，并合理分配资源，从而在部署更高复杂度的模型之前有效筛除表现不佳的特征映射。因此，特征映射选择的计算负担大幅降低：其成本从全连接核训练的多项式复杂度（通常为 $\mathcal{O}(N^3)$）降至由采样轴数量 t 主导的可扩展成本，同时 t 可以保持较小而仍维持严格的统计置信度。

本文其余部分安排如下。第 2 节构建理论框架，正式定义轴对齐回归评分，并证明其构成最优训练误差的认证上界。第3 节介绍旨在克服特征空间指数级扩展问题的蒙特卡洛估计策略；本节还推导了严格的非渐近统计保证，并详细说明用于样本量校准的自适应算法。实验方法，包括数据集生成和特征映射配置，将在第 4节详述。第 5 节展示数值基准测试，将估计量与经过充分训练的量子和经典回归器进行比较，以验证其紧致性和预测能力。最后，第 6 节总结全文，并展望未来研究方向。

2 量子回归的泡利轴上界

我们考虑量子核框架中的监督回归，其目标是根据数据集$D = \{(\mathbf{x}_k, y_k)\}_{k=1}^N$，从输入 $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^m$ 预测实值标签 $y \in \mathbb{R}$ 。在这种方法中，经典数据通过量子特征映射 $|\Phi(\mathbf{x})\rangle = U(\mathbf{x})\,|0\rangle^{\otimes n}$嵌入到一个 n量子比特的希尔伯特空间中，其中 $U(\mathbf{x})$ 是参数化编码电路。两个编码输入之间的相似性通过保真度核

进行量化

$$K(\mathbf{x}, \mathbf{z}) = \left|\langle \Phi(\mathbf{x}) \mid \Phi(\mathbf{z}) \rangle\right|^2, \tag{1}$$

可通过相应量子电路的跃迁概率在量子设备上进行估计 [5], 而回归模型本身则通过经典方法进行优化。

为分析这一映射的几何结构，我们考虑相应的密度算符 $\rho(\mathbf{x}) = |\Phi(\mathbf{x})\rangle\langle\Phi(\mathbf{x})|$。虽然底层的 n量子比特希尔伯特空间具有复维数 2^n，但作用于其上的厄米算符构成一个实向量空间。该算符空间自然由 n量子比特泡利算符集合 $\mathcal{P}_n = \{I, X, Y, Z\}^{\otimes n}$ [20] 张成。因此，任意纯态密度矩阵都可以在此基底下展开为 $\rho(\mathbf{x}) = \frac{1}{2^n} \sum_{i=1}^{d} a_i(\mathbf{x})\, \sigma^i$，其中 $\sigma^i \in \mathcal{P}_n$ 且 $a_i(\mathbf{x}) = \mathrm{tr}[\rho(\mathbf{x})\ \sigma^i]$ 是第 i个泡利算符的期望值。利用正交关系 $\mathrm{tr}(\sigma^i \sigma^j) = 2^n \delta_{ij}$，量子核可以改写为

$$K(\mathbf{x}, \mathbf{z}) = \frac{1}{2^n} \sum_{i=1}^{d} a_i(\mathbf{x})\, a_i(\mathbf{z}), \tag{2}$$

其中， d 表示泡利特征空间的总维度；通常，在考虑全连接泡利基时， $d = 4^n$ 。这种重新表述揭示了量子核等价于维度为 4^n的实特征空间中的标准线性核，为量子描述与经典描述搭建了直接的几何桥梁。

2.1 轴对齐最小二乘上界

基于这一表示，我们现在对最小二乘回归模型进行形式化。每个编码后的输入$\mathbf{x}_k$ 都映射为一个特征向量$\mathbf{a}(\mathbf{x}_k) = [a_1(\mathbf{x}_k) \ldots, a_d(\mathbf{x}_k)]$。给定一组目标标签 $\{y_k\}_{k=1}^N$，任何预测器$f : \mathcal{X} \to \mathbb{R}$ 的质量均按照经验风险最小化的标准框架 [2]，以经验均方误差衡量：

$$\mathrm{MSE}(f) = \frac{1}{N} \sum_{k=1}^{N} \big(y_k - f(\mathbf{x}_k)\big)^2. \tag{3}$$

我们将泡利特征空间中的全连接仿射线性回归器族定义为$\mathcal{F} = \{f(\mathbf{x}) = \langle \mathbf{w},\ \mathbf{a}(\mathbf{x}) \rangle + b \mid \mathbf{w} \in \mathbb{R}^d,\ \ b \in \mathbb{R}\}$。理想情况下，我们希望在 $\mathcal{F}$中找到使式（3）最小的预测器，从而达到最优训练误差均方误差$^* = \min_{f \in \mathcal{F}} \mathrm{MSE}(f)$。然而，要刻画这一最小值，通常需要完全访问规模呈指数级增长的特征空间。作为一种可行的替代方案，我们考虑轴受限预测器这一子类，它们仅依赖于单个泡利特征 $i \in \{1, \ldots, d\}$:

$$\mathcal{F}_{\text{axis}}^{(i)} = \Big\{ f_i(\mathbf{x}) = w\, a_i(\mathbf{x}) + b \;\Big|\; w, b \in \mathbb{R} \Big\}. \tag{4}$$

我们将仅利用第 i个轴所能达到的最小误差记为均方误差$_i$ $=\min^{f\in\mathcal{F}^{(i)}\ \text{轴}}\mathrm{MSE}(f)$，并将单轴表现的整体最优值记为均方误差$_{\text{axis}}$ $=\min_i \mathrm{MSE}_i$。

这些假设类的嵌套结构直接带来了可证明的性能保证。

定理 1（泡利轴上界） *For any dataset D encoded by a fixed quantum featuremap, the optimal affine linear regression error is upper bounded by the best axis-aligned error:*

$$\mathrm{MSE}^* \le \mathrm{MSE}_{\text{axis}}.$$

Proof 由于对于每个坐标 $\mathcal{F}^{(i)}$ 轴$\subset \mathcal{F}$ i，较大集合 $\mathcal{F}$ 上的最小值不会超过子集 $\mathcal{F}^{(i)}$ 轴上的最小值。因此，对于所有 i，均方误差$^* \le$ 均方误差$_i$。对所有 d 坐标取最小值，可得均方误差$^* \le \min_i$均方误差$_i =$ 均方误差$_{\text{axis}}$。□

这一结果可自然推广到更复杂的假设类 $\mathcal{H} \supseteq \mathcal{F}$，例如在诱导特征空间上使用核岭回归或支持向量回归（SVR）时所采用的假设类。由于这些方法在包含所有线性函数的空间上进行优化，其经验误差满足 $\inf_{h\in\mathcal{H}}$均方误差$(h) \le$ 均方误差*。因此，均方误差$_{\text{axis}}$ 可作为一种保守且无需训练的代理指标：如果简单的轴对齐模型取得较低误差，那么从数学上可以保证，表达能力更强的核模型至少同样出色。不过，要将这一上界用作诊断工具，我们需要一个以可观测数据统计量表示的、明确且可计算的均方误差$_{\text{axis}}$ 表达式。

2.2 解析推导与复杂度

现在，我们推导定理 1 中界的显式解析表达式。对于任意固定的泡利坐标 ，最优预测器 轴 对应于一元普通最小二乘法（普通最小二乘法）解。其参数仅取决于数据的经验矩。下面定义样本均值、方差和协方差：

$$\bar{y} = \frac{1}{N}\sum_{k=1}^{N} y_k, \qquad \bar{a}_i = \frac{1}{N}\sum_{k=1}^{N} a_i(\mathbf{x}_k),$$
$$\mathrm{Var}(a_i) = \frac{1}{N}\sum_{k=1}^{N}\big(a_i(\mathbf{x}_k) - \bar{a}_i\big)^2, \quad \mathrm{Cov}(a_i, y) = \frac{1}{N}\sum_{k=1}^{N}\big(a_i(\mathbf{x}_k) - \bar{a}_i\big)\big(y_k - \bar{y}\big). \tag{5}$$

如果 $\mathrm{Var}(a_i) > 0$，则最优系数 $w_i^\star$和 $b_i^\star$由标准普通最小二乘法结果给出：

$$w_i^\star = \frac{\mathrm{Cov}(a_i, y)}{\mathrm{Var}(a_i)}, \qquad b_i^\star = \bar{y} - w_i^\star \bar{a}_i. \tag{6}$$

将这些代入误差函数，可得残差方差：

$$\mathrm{MSE}_i = \mathrm{Var}(y) - \frac{\mathrm{Cov}(a_i, y)^2}{\mathrm{Var}(a_i)} = \mathrm{Var}(y)\big(1 - \rho^2_{y,a_i}\big), \tag{7}$$

其中 $\rho_{y,\ a_i}$ 是目标变量 y 与特征a_i之间的皮尔逊相关系数。该等式源于简单线性回归的一项基本性质：决定系数 R^2 恰好等于相关系数 [21]的平方。在 $\mathrm{Var}(a_i) = 0$ 的退化情形下，预测器退化为常数均值 $f_i(\mathbf{x}) = \bar{y}$，从而 $\mathrm{MSE}_i = \mathrm{Var}(y)$。

将这一结果代回定理 1，即可得到全局上界的闭式表达式：

$$\mathrm{MSE}_{\mathrm{axis}} = \min_{1\le i\le d} \mathrm{MSE}_i = \mathrm{Var}(y)\left(1 - \max_{1\le i\le d} \rho^2_{y,a_i}\right). \tag{8}$$

式 (8) 提供了有力的几何解释：经认证的上界完全由与目标变量边际相关性最强的单个泡利轴决定。

关于计算复杂度，计算式 (8) 需要遍历所有特征维度。理想情况下，时间复杂度为 $O(Nd)$，空间复杂度为 $O(d)$。不过，实际应用中需要考虑两个因素。首先，当通过有限次量子测量估计泡利期望值时，经验矩会受到抽样噪声的影响。从统计学角度看，这会在回归变量中引入测量误差，从而导致衰减偏差 [22]：噪声会人为增大 $\mathrm{Var}(a_i)$，并系统性地缩小 $|\rho_{y,\ a_i}|$。因此，计算得到的 $\mathrm{MSE}_{\mathrm{axis}}$ 会偏保守（大于理想无噪声情况下的值），但仍能保证上界有效。

其次，更关键的是，维度 d对应于全连接算符基的大小，即 $d = 4^n$。虽然计算复杂度与 d呈线性关系，但随着量子比特数量增加，算符基会呈指数增长，因此对于多量子比特系统，精确计算 $\max_i \rho^2_{y,\ a_i}$ 将难以实现。这一扩展性瓶颈促使我们在下一节引入蒙特卡罗估计框架。

3 蒙特卡罗估计框架

为克服第 2 节中指出的泡利基（$d = 4^n$）的指数扩展问题，我们引入一种概率估计方法。近期已有研究提出类似的蒙特卡罗策略，用于估计量子分类器的最低准确率 [23]。本文将这一框架扩展并加以形式化，使其适用于连续回归领域，并推导出均方误差（MSE）的特定集中界。

我们不再扫描所有 d坐标，而是从轴中均匀采样一个包含 t 个轴的随机子集 $T \subset \{1, \ldots, d\}$，其中 $t \ll d$。我们将该子集中的最小误差定义为蒙特卡洛估计量：

$$\widehat{\mathrm{MSE}}_{\mathrm{axis}}(T) := \min_{i\in T} \mathrm{MSE}_i = \mathrm{Var}(y)\left(1 - \max_{i\in T} \rho^2_{y,a_i}\right). \tag{9}$$

该估计量从子集 T 中选出最佳轴，从而为均方误差$_{\text{axis}}$提供一种可高效计算的替代量。关键在于，尽管这是一个近似值，该估计量仍保留了精确度量的严格性能保证，正如以下定理所证明的那样。

定理 2（蒙特卡洛界） *For any non-empty subset* $T \subseteq \{1, \ldots, d\}$, *the optimal affine linear error* 均方误差* *is upper bounded by the Monte Carlo estimator:*

$$\text{MSE}^* \leq \text{MSE}_{\text{axis}} \leq \widehat{\text{MSE}}_{\text{axis}}(T). \tag{10}$$

Proof 根据定义，均方误差$_{\text{axis}}$ = 最小值$_{1\leq j\leq d}$ 均方误差$_j$。由于子集上的最小值 T必然大于或等于全集上的最小值（即 $T \subseteq \{1, \ldots, d\}$ $\Longrightarrow$最小值$_{i\in T}\, x_i$ $\geq$ 最小值$_{all}\, x_i$），因此均方误差$_{\text{axis}}$ $\leq$ $\widehat{\text{均方误差}}_{\text{axis}}$（$T$）。结合定理1所述的均方误差* $\leq$ 均方误差$_{\text{axis}}$，可得不等式链：均方误差* $\leq$均方误差$_{\text{axis}}$ $\leq$ $\widehat{\text{均方误差}}_{\text{axis}}$（$T$）。□

式（10）中的不等式证实，蒙特卡洛估计量是最优误差均方误差的经认证上界*，且无需模型训练。此外，这个$\widehat{\text{}}$上界可以系统地收紧：映射 $T \mapsto$ 均方误差$_{\text{axis}}$（T）关于集合包含关系单调不增。具体而言，对于嵌套子集 $T_t \subset T_{t+1}$，估计$\widehat{\text{}}$量满足均方误差$_{\text{axis}}$（T_{t+1}） $\leq$均方误差$_{\text{axis}}$（T_t），并收敛至精确值均方误差$_{\text{axis}}$，当$t \to d$。

3.1 统计保证与阈值选择

$\widehat{\text{}}$虽然定理 2 确立了蒙特卡洛估计量的均方误差$_{\text{axis}}(T)$ 是最优仿射回归误差的有效上界，但它并未量化在实践中该上界足够紧的概率。在回归任务中，紧密程度必须相对于目标变量的内在尺度来解释。因此，我们引入目标阈值 τ，其定义基于期望的决定系数$R^2_{\text{target}} \in (0, 1]$:

$$\tau = \text{Var}(y)\big(1 - R^2_{\text{target}}\big). \tag{11}$$

等价地，我们可以写成 $\tau = \tau_{\text{ratio}} \cdot \text{Var}(y)$，其中 $\tau_{\text{ratio}} := 1 - R^2_{\text{target}}$。

任何满足均方误差 i 的轴$_i \leq \tau$ 因此至少解释了样本方差的 R^2_{target}。为了刻画此类轴在泡利特征空间中的出现频率，我们定义优良轴的集合

$$\mathcal{A}_\tau = \{i \in \{1, \ldots, d\} : \text{MSE}_i \leq \tau\}, \tag{12}$$

并在有限轴集合上引入经验累积分布函数，

$$F(\tau) = \frac{1}{d}\,|\mathcal{A}_\tau|. \tag{13}$$

我们用 $p := F(\tau)$ 表示达到目标性能水平的轴所占的比例。

以下结果量化了随机轴子集至少包含一个良好轴的概率。

定理 3 *Fix a threshold τ and let $p = F(\tau)$. If $T \subset \{1, \dots, d\}$ is sampled uniformly without replacement with $|T| = t$, then*

$$\mathbb{P}\Big(\widehat{\mathrm{MSE}}_{\mathrm{axis}}(T) \le \tau\Big) = 1 - \frac{\binom{d-\lfloor pd \rfloor}{t}}{\binom{d}{t}} \;\ge\; 1 - (1-p)^t. \tag{14}$$

Proof 令 $\mathcal{A}_\tau$ 表示优良轴的集合，其基数为 $m = |\mathcal{A}_\tau| = \lfloor pd \rfloor$。事件$\widehat{\text{均方误差}}_{\mathrm{axis}}(T) \le \tau$当且仅当 $T \cap \mathcal{A}_\tau \neq \emptyset$时发生。在不放回均匀抽样下，$T$避开 $\mathcal{A}_\tau$ 的概率由超几何项 $\binom{d-m}{t}/\binom{d}{t}$给出。不等式由界 $\binom{d-m}{t}/\binom{d}{t} \le (1-m/d)^t = (1-p)^t$推出。□

定理 3 的一个直接推论是，给出达到所需置信水平所需样本量的充分条件。

推论 4 *Fix $\delta \in (0,1)$ and suppose that at least a fraction $p \in (0,1)$ of the axes satisfies*均方误差 *$_i \le \tau$. If T is sampled uniformly without replacement with $|T| = t$, then the condition*

$$t \;\ge\; \frac{\log(1/\delta)}{\log\big(1/(1-p)\big)} \tag{15}$$

ensures

$$\mathbb{P}\Big(\widehat{\mathrm{MSE}}_{\mathrm{axis}}(T) \le \tau\Big) \ge 1 - \delta.$$

In particular, since $-\log(1-p) \ge p$ for all $p \in (0,1)$ the simpler sufficient condition

$$t \;\ge\; \frac{1}{p} \log\frac{1}{\delta} \tag{16}$$

also guarantees the same confidence level.

*Proof*根据定理 3，$\mathbb{P}(\widehat{\text{均方误差}}_{\mathrm{axis}}(T) \le \tau) \ge 1-(1-p)^t$。要求右侧至少为 $1-\delta$，可得 $(1-p)^t \le \delta$，这等价于 (15)。由 16) 可得界 $-\log(1-p) \ge p$。□

3.2 自适应样本量校准

推论 4 中的样本量界取决于满足阈值条件的轴所占的未知比例 $p = F(\tau)$。由于通过穷举扫描 p 来计算 $d = 4^n$ 泡利轴在规模较大时难以实现，我们采用自适应蒙特卡洛过程在线估计该量，同时保持严格的概率保证。

固定阈值 τ 后，每个抽取的轴 i 都对应一个指示变量

$$X_i := \mathbf{1}\{\mathrm{MSE}_i \le \tau\}, \tag{17}$$

该变量记录轴是否达到目标性能。尽管抽样是从有限总体中不放回进行的，我们仍采用单侧 Hoeffding 界，这种界在此情形下依然保守。

对 t 个不同轴中的子集 T 进行抽样后，定义

$$s := \sum_{i \in T} X_i, \qquad \hat{p} := \frac{s}{t}, \tag{18}$$

为成功轴的数量及相应的经验成功率。对于置信参数 $\alpha \in (0,1)$，p 的置信下界为

$$p_L := \max\left\{0,\ \hat{p} - \sqrt{\frac{1}{2t}\log\frac{1}{\alpha}}\right\}, \tag{19}$$

其成立的概率至少为 $1-\alpha$。

将 p_L 代入定理 3 的精确覆盖条件，可得到所需样本量的数据驱动估计，

$$t_{\mathrm{req}} := \frac{\log(1/\delta)}{\log\!\big(1/(1-p_L)\big)}, \tag{20}$$

约定当 $p_L = 0$ 时 $t_{\mathrm{req}} = \infty$。迭代抽样直至满足停止条件 $t \geq t_{\mathrm{req}}$，此时估计量满足$\mathbb{P}(\text{均方误差}_{\mathrm{axis}}(T) \leq \tau) \geq 1-\delta$。

为避免在高质量轴极其稀疏的情形下执行时间无上限，我们设置了最大采样预算 $t_{\max}$。此外，还采用无效停止准则：如果未观察到合格轴（$\hat{p} = 0$），且置信区间宽度低于最小容差，则提前终止该过程。在这种情况下，$\widehat{\text{返回值均方误差}}_{\mathrm{axis}}(T)$ 依据定理 2，仍是最优回归误差的有效认证上界，但不再保证满足阈值。该自适应策略总结于算法 1，将理论覆盖界转化为实用且稳健的停止规则，并根据高性能轴的经验密度分配计算资源。

算法1 形式化定义了实验部分中始终使用的自适应蒙特卡洛过程，并给出了经过认证的停止规则，从而避免对 4^n 轴对齐假设空间进行穷举扫描。

4 实验设置

为验证理论框架和自适应蒙特卡洛过程，我们设计了一套全面的实验套件，考察结构复杂度、噪声水平和特征映射架构的变化。

算法 1 轴对齐界的自适应蒙特卡罗校准

要求： τ_{ratio}, δ_{total}, t_0, b, $t_{\max}$, $\epsilon_{\min}$

确保： 抽样集 T，估计量 $\widehat{\text{均方误差}}_{\text{axis}}(T)$

$\tau \leftarrow \tau_{\text{ratio}}$方差$(y)$, $\alpha = \delta = \delta_{\text{total}}/2$

抽取初始集 T ($|T| = t_0$); $t \leftarrow t_0$, $s \leftarrow \sum_{i \in T} \mathbf{1}\{\text{均方误差}_i \le \tau\}$

</style}**while** $t < t_{\max}$ **do**

$\hat{p} \leftarrow s/t$, $\epsilon \leftarrow \sqrt{\frac{1}{2t}\text{对数}\frac{1}{\alpha}}$, $p_L \leftarrow \max\{0, \hat{p} - \epsilon\}$

$t_{\text{req}} \leftarrow$ 对数$(1/\delta)$/对数$(1/(1 - p_L))$

如果 $t \ge t_{\text{req}}$ **或** $(\hat{p} = 0 \wedge \epsilon < \epsilon_{\min})$ **则跳出循环**

结束如果

用新轴更新 T b ; $s \leftarrow s + \Delta s$, $t \leftarrow t + b$

结束循环

返回 T , $\min_{i \in T} \text{MSE}_i$

4.1 数据集与基准测试策略

为严格验证自适应估计器的表现，我们采用混合基准测试策略。首先，我们构建了两个几何特性截然相反的合成景观：一个稠密且具有相关性，另一个稀疏且高维，用作受控压力测试。由于这些任务的真实结构是预先已知的，因此我们可以验证估计器的收敛行为是否符合理论预期。最后，我们在一个真实世界数据集上检验该方法，以评估其在非结构化噪声下的稳健性。

表 1总结了三种具有代表性的回归任务：(i)合成相关高斯，代表具有相关特征的中等情形；(ii)合成稀疏，代表具有高维稀疏依赖的困难情形；以及 (iii) 真实加利福尼亚，即加利福尼亚住房数据集的一个子集，降至 $n = 4$ 个主成分，作为带有自然噪声的真实世界基准。

4.2 量子特征映射与配置空间

量子特征空间由泡利特征映射拟设 [5], 生成，该拟设通过交错的 Hadamard 层和依赖于数据的纠缠门，将经典数据 $\mathbf{x} \in \mathbb{R}^n$ 映射为一个 n 量子比特态:

$$\mathcal{U}_{\Phi(\mathbf{x})} = \left(U_{\Phi(\mathbf{x})} H^{\otimes n}\right)^r, \qquad U_{\Phi(\mathbf{x})} = \exp\left(i \sum_{S \in \mathcal{I}} \phi_S(\mathbf{x}) \prod_{k \in S} P_k\right). \tag{21}$$

其中， $P_k \in \{X, Y, Z\}$ 是泡利算符， $\mathcal{I}$ 定义了连通图，而 $\phi_S(\mathbf{x})$ 是经典编码函数。

为确保稳健性和架构多样性，我们的研究不依赖单一模型，而是对总计 324 种实验配置进行了全面评估，其中针对每个数据集应用了 108 种不同的特征映射。我们通过排列表 2 中定义的三个关键组成部分，系统地探索这一设计空间：(1)经典预处理，用于处理数据缩放；(2) 泡利

Ta表 1 数值实验所用数据集概况　　验。

属性	详情
数据集	`synthetic_corr_gauss`
情形	**中等（密集）：** 信息分布在相关轴上；近最优解的密度较高。
目标函数	$y = \exp(-\frac{1}{2}\mathbf{x}^\top \Sigma^{-1}\mathbf{x})$，其中 $\Sigma_{ij} = 0.3 + 0.7\delta_{ij}$。
数据集	`synthetic_sparse`
情形	**困难（稀疏）：** 信息集中在隐藏子集；信息轴密度较低（大海捞针）。
目标函数	$y = \sum_{i\in\mathcal{S}} \sin(x_i) + 0.1\sum_{j\notin\mathcal{S}} x_j$，其中 $\mathcal{S}$ 是活跃子集。
数据集	`real_california`
情形	**真实世界：** 具有非结构化噪声和复杂依赖关系的自然数据分布关系。
描述	加利福尼亚住房（经主成分分析降至 $n = 4$，保留 95% 的方差）。

序列，用于确定量子关联的基础；以及 (3) 数据映射规则，用于控制相位编码的非线性。我们进一步将这些核心组成部分与不同的纠缠模式和电路深度 ($r \in \{1, 2\}$) 结合，以涵盖广泛的模型复杂度。

表 2 配置空间定义。泡利序列包括 `all pairs`，其中包含所有双体组合（`XY`、`YZ`、`ZZ` 等）。

阶段	选项和定义	定性效果
经典预处理	`id`: $f(x) = x$ `tanh`: $f(x) = \tanh(x)$ `rbf-s1`: $f(x) = e^{-x^2/2}$	无界输入（原始数据） 限制在 $[-1, 1]$，抗离群值 局部相似性强调（高斯）
泡利序列	`Z+ZZ`（[Z, ZZ]） `Y+YY`([Y, YY]) `all pairs`（[XY、YZ、ZZ.....]）	计算基关联 非对角干涉（复数） 更丰富的纠缠结构
数据映射 ($\phi(x)$)	`prod`: $\prod_i x_i$ `pi*prod`: $\pi\prod_i x_i$ `sum+prod`: $\sum_i x_i + \prod_i x_i$	纯非线性交叉项 全连接 2π 相位覆盖 加性混合及乘性项

5 数值性能与基准测试

本节通过将认证界与全连接训练回归模型的性能进行比较，评估其预测能力。我们的主要目标是

表 3 锚点配置：对比无需训练的蒙特卡洛界（均方误差$_{\text{axis}}$）、精确单轴最优值（均方误差$_{\text{axis}}$）和完全训练的模型。

数据集	配置	均方误差$_{\text{axis}}$	$\widehat{\text{均方误差}}_{\text{axis}}$	t_{used}	均方误差$_{\text{ridge}}$	均方误差$_{\text{svr}}$
真实加利福尼亚	rbf-s1 | 乘积 | Z+ZZ | 线性 | r=2	0.6863	0.6863	110	0.4123	0.5105
合成相关高斯数据集	rbf-s1 | 乘积 | Y+YY | 线性 | r=1	0.3195	0.7053	50	0.1165	0.0101
syn_sparse	双曲正切 | 乘积 | Z+ZZ | 全连接 | r=1	0.2684	0.2684	256*	0.0000	0.0094

*因预算限制而停止。

证明免训练估计器的均方误差$_{\text{axis}}(T)$能够有效识别高质量特征映射，而无需进行代价高昂的变分优化。

对于 $n = 4$ 量子比特（$d = 256$），我们首先在每个数据集的108种配置网格上进行了无需训练的扫描。自适应估计器的置信水平设置为 $\delta_{\text{total}} = 0.05$ ，相对阈值设置为 $\tau_{\text{ratio}} = 0.95$。采样以试运行样本量 $t_0 = 50$ 初始化，并以 $b = 20$为批次进行更新，最大预算为 $t_{\max} = d$。

对于每个数据集，我们仅根据自适应估计器确定了表现最佳的单一架构，称为“锚点”配置。为验证这一选择，随后我们专门针对这些锚点训练了两个回归模型：

1. **量子岭回归（**均方误差$_{\text{ridge}}$**）：** 一种带有 ℓ_2 正则化（$\alpha = 10^{-3}$）的线性回归器，直接在 d维特征向量 $\Phi(\mathbf{x})$ 上训练。它可作为无噪声量子支持向量回归器（QSVR）的理想代理。2. **经典支持向量回归（**均方误差$_{\text{svr}}$**）：** 一种在原始经典数据上训练的标准支持向量回归器，采用径向基函数核，作为基线。

5.1 结果与讨论

表3总结了针对每个数据集确定的最佳“锚点”配置的结果（数据集概况见表1 ）。结果揭示了认证界、数据集几何结构与实际训练潜力之间的对齐关系的三个关键见解：

1. 真实世界复杂性与线性组合。

对于真实加利福尼亚，估计量高效地收敛了（$t = 110 < d/2$），并达到了穷举界。然而，最佳单轴特征（MSE$_{\text{axis}} \approx 0.68$）与经过完整训练的量子岭回归模型（MSE$_{\text{ridge}} \approx 0.41$）之间仍存在显著差距。这种表现符合“真实世界”情形的描述：尽管自适应过程正确识别出了最佳的单个泡利特征，但训练后的回归器表现更优，这证实了解决现实任务需要将多个特征线性组合（叠加），而非仅使用一个最优轴。值得注意的是，量子岭回归的表现优于经典

支持向量回归（0.51），这表明Pauli特征映射比标准径向基函数核更能捕捉房价数据中的非平凡相关性。

2. 稀疏情形下的特征映射对齐。
在 `syn sparse` 情形下，岭回归器实现了完美重建（均方误差 ≈0.0000），从理论上验证了所选特征映射（`Z+ZZ`）能够完美张成数据集的生成函数。自适应估计器在此提供了关键的诊断信息：尽管它耗尽了预算（$t = 256$），证实了这是“困难/稀疏”情形——优良轴如同大海捞针般罕见——但它仍成功找到了最优轴（差距 $= 0.0$）。这证明了该方法的稳健性：即使稀疏性导致统计认证十分困难，搜索过程仍能有效找到适用于优化的高质量特征。

3. 相关数据中的“稠密陷阱”。
`syn corr gauss` 数据集展示了“中等/稠密”情形的微妙之处。估计量提前停止（$t = 50$），且给出的界较松。这并非失败，而是因为在相关景观中，“相当不错”的轴密度很高。算法很快满足阈值条件 τ 并按设计停止，尚未遇到罕见得多的全局最优解。此外，经典支持向量回归（0.0101）在此任务中表现更优，这符合预期，因为目标变量是平滑的高斯函数，与经典径向基函数核天然契合，却更难用离散的泡利串近似（岭回归器 ≈ 0.11）。

总而言之，蒙特卡洛估计量是一种经济高效的探测工具：它正确识别出`syn sparse`的稀疏性（通过预算耗尽），以及`syn corr gauss` 的稠密性（通过提前停止），同时为真实场景提供了紧致的预测上界。

6 结论

在这项工作中，我们为量子核回归的最优训练误差提出了一个经过认证且无需训练的上界。通过将 Suzuki等人提出的轴对齐框架从分类扩展到回归问题，我们推导出一个闭式度量，将泡利特征空间的几何结构与均方误差直接关联起来。为应对特征维度的指数级增长，我们提出了一种自适应蒙特卡洛估计量，并基于 Hoeffding集中不等式为其提供严格的统计保证。

在合成数据集和真实世界数据集上进行的大量数值实验，证实了该方法的效率和预测能力。无论是在稀疏景观还是稠密景观中，估计量都能稳定收敛到真实的穷举界，且所需采样预算远小于全基底维度。此外，与经过完整训练的岭回归器进行比较后发现，认证界并非仅仅是宽松的理论上限，而是能够高保真地预测给定特征映射的表达能力，从而无需承担基于梯度的优化成本，就能可靠地区分性能较高和较低的架构。

尽管认证过程本质上依赖于阈值，但这种依赖应被视为一种特性，而非局限。在我们的实验中，为确保在不同数据集之间进行公平比较，我们采用了固定的超参数；然而，阈值比率 τ_{ratio} 和置信水平 δ_{total} 等参数，可为选择性与采样成本之间的权衡提供有原则的调控手段。这种灵活性使实践者能够根据数据集的特定特征和可用计算资源调整认证流程。

总体而言，所提出的框架为量子模型选择在 NISQ时代建立了一种实用的诊断工具。通过在训练前筛除效果不佳的特征映射配置，该框架能够将计算资源专门用于最有前景的量子架构。未来一个自然的研究方向是，将自适应过程识别出的信息轴子集用作压缩特征空间，从而在保持较低计算开销的同时，为变分优化提供指导或加以约束。

参考文献

[1] Russell, S.J., Norvig, P.：《人工智能：一种现代方法》，第4版。培生人工智能系列。培生，波士顿，马萨诸塞州（2020）[2] Bishop, C.M.：《模式识别与机器学习》。信息科学与统计学。施普林格，纽约，纽约州（2006）[3] Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., Lloyd, S.：量子机器学习。《自然》**549**（7671），195–202（2017）[4] Grant, E., Benedetti, M., Cao, S., Hallam,A., Lockhart, J., Stojevic, V., Green, A.G., Severini, S.：层次量子分类器。npj 量子信息**4**（1），65（2018）[5] Havlíˇcek, V., Córcoles, A.D., Temme, K., Harrow, A.W., Kandala, A., Chow, J.M., Gambetta, J.M.：利用量子增强特征空间进行监督学习。《自然》**567**（7747），209–212（2019）[6] Schuld, M., Killoran, N.：特征希尔伯特空间中的量子机器学习。《物理评论快报》**122**（4），040504（2019）[7]Mitarai, K., Negoro, M., Kitagawa, M., Fujii, K.：量子电路学习。《物理评论A辑》**98**（3），032309（2018） https://doi.org/10.1103/PhysRevA.98.032309[8] Huang, H.-Y., Broughton, M., Mohseni, M., Babbush, R., Boixo, S., Neven, H., McClean, J.R.：量子机器学习中的数据力量。《自然通讯》**12**（1），2631（2021）[9] Li, M.-C., Liu, J.-G., Pan, X.-Y., Fan, H.：利用随机特征映射在NISQ设备上评估量子核。《量子科学与技术》**7**（4），

[10] Tacchino, F., Macchiavello, C., Gerace, D., Bajoni：人工神经网络的量子实现。npj 量子信息 5（1），26（2019）[11] Gentinetta, G., Thomsen, A., Sutter, D., Woerner, S.：量子支持向量机的复杂性。Quantum 8，1225（2024）https://doi.org/10.22331/ q-2024-01-11-1225[12] McClean, J.R., Boixo, S., Smelyanskiy, V.N., Babbush, R., Neven, H.：量子神经网络训练景观中的贫瘠高原。《自然通讯》9（1），4812（2018）https://doi.org/10.1038/s41467-018-07090-4[13] Cerezo, M., Sone, A., Volkoff, T., Cincio, L., Coles, P.J.：浅层参数化量子电路中依赖于代价函数的贫瘠高原。《自然通讯》12（1），1791（2021）https://doi.org/10.1038/s41467-021-21728-w [14] Larocca, M., Cerezo, M., Sharma, K., Sone, A., Coles, P.J., Holmes, Z.：变分量子计算中的贫瘠高原。《现代物理评论》（2024）。预印本见 arXiv:2205.05756[15] Sim, S., Johnson, P.D., Aspuru-Guzik, A.：量子-经典混合算法中参数化量子电路的表达能力与纠缠能力。《先进量子技术》2（12），1900070（2019）https://doi.org/10.1002/ qute.201900070[16] Preskill, J.：NISQ时代及之后的量子计算。Quantum 2，79（2018）https://doi.org/10.22331/q-2018-08-06-79[17] Smola, A.J., Sch¨olkopf, B.：支持向量回归教程。《统计与计算》14（3），199–222（2004）https://doi.org/10.1023/B:STCO.0000035301. 49549.88[18] Suzuki, Y., Yano, H., Gao, Q., Uno, S., Tanaka, T., Akiyama, M., Yamamoto, N.：基于核的量子分类器的特征映射分析与综合。《量子机器智能》4（1），1–15（2022）https://doi.org/10.1007/ s42484-022-00058-8[19] Hoeffding, W.：有界随机变量之和的概率不等式。《美国统计学会杂志》58（301），13–30（1963）https://doi.org/10.1080/01621459.1963.10500830[20] Nielsen, M.A., Chuang, I.L.：《量子计算与量子信息：十周年纪念版》。剑桥大学出版社，剑桥（2010）[21]Kutner, M.H.：《应用线性统计模型》。麦格劳-希尔国际版。麦格劳-希尔欧文，纽约（2005）

[22]Fuller, W.A.: 《测量误差模型》。威利概率与统计丛书。约翰·威利父子公司，纽约（1987）。https://doi.org/10.1002/9780470316665

[23] Gonçalves, D.N.,Fernandes, T.D., Cordeiro, A.M.M., Lugao, P.H.G., Dias, J.T.: 量子核方法中最低准确率的经认证下界与高效估计（2025）。https://arxiv.org/abs/2512.20588