arXiv:2602.07141v1 [cs.LG] 6 Feb 2026

用于学习和神经网络的特征再生核巴拿赫空间

Isabel de la Higuera[1], Francisco Herrera[1,2],和 M. Victoria Velasco[3,4]

[1]计算机科学与人工智能系，安达卢西亚数据科学与计算智能研究所 (DaSCI)，格拉纳达大学，西班牙格拉纳达，18140，电子邮箱：herrera@decsai.ugr.es [2]ADIA实验室，阿布扎比，阿拉伯联合酋长国。[3]数学分析系，理学院，格拉纳达大学，西班牙格拉纳达 18071，电子邮箱：vvelasco@ugr.es[4]通讯作者

2026年2月10日

摘要

再生核希尔伯特空间为基于核的学习提供了基础框架，在这一框架下，正则化和插值问题可通过经典表示定理获得有限维解。然而，许多现代学习模型——包括配备非二次范数的固定架构神经网络——自然会产生非希尔伯特几何结构，因而超出这一框架的适用范围。在巴拿赫空间中，仅有点求值泛函的连续性并不足以保证特征表示或基于核的学习表述。

本文基于特征再生核巴拿赫空间的概念，构建了一个用于巴拿赫空间中学习的泛函分析框架。我们明确了在希尔伯特空间情形之外，特征映射、核构造和表示定理类结果得以恢复所需的精确结构条件。在这一框架下，监督学习被表述为最小范数插值问题或正则化问题，并建立了存在性结果以及条件表示定理。

我们进一步将该理论扩展到向量值特征再生核巴拿赫空间，并证明固定架构神经网络自然会导出这类空间的特殊实例。这为核方法和神经网络提供了统一的函数空间视角，并阐明了基于核的学习原则何时能够扩展到再生核希尔伯特空间之外。

关键词
再生核希尔伯特空间、再生核巴拿赫空间、特征再生核巴拿赫空间、特征映射、核函数、特殊特征再生核巴拿赫空间、向量值特征再生核巴拿赫空间、最小范数插值问题、正则化问题、表示定理、神经网络。

1 引言

机器学习的数学基础深深植根于泛函分析，尤其是再生核希尔伯特空间（RKHSs）理论。自 Aronszajn（1950）的奠基性工作（Aronszajn，1950）以来，再生核希尔伯特空间理论为从有限数据集中学习提供了严谨且统一的框架，使支持向量机、核岭回归和高斯过程等经典算法能够被表述为泛函优化问题。在这一框架下，学习自然对应于赋有再生核的函数希尔伯特空间中的最小范数插值或正则化。此外，点求值泛函的连续性等价于再生核的存在性，进而也意味着存在相应的特征映射；这又为广泛的范数单调正则化学习问题中的经典表示定理提供了基础（Schölkopf &Smola，2002）。

然而，许多与现代机器学习相关的函数类——尤其是配备非二次参数范数的固定架构神经网络——并不自然地适合用希尔伯特空间来表述（Zhang, Xu,& Zhang，2009）。在巴拿赫空间中，仅有点求值的连续性并不足以保证特征表示或表示定理的存在，还需要额外的结构性假设，正如我们将在本文中所展示的那样。受这一空白的启发，再生核巴拿赫空间（RKBSs）被提出作为再生核希尔伯特空间的推广；然而，在何种精确条件下能够恢复核构造和表示定理类结果，目前仍未得到充分理解。

在这项工作中，我们确定了在巴拿赫空间中实现特征映射诱导的再生核巴拿赫空间及条件表示定理所需的结构要求，并证明固定的神经网络架构会自然地诱导出特殊的向量值特征再生核巴拿赫空间。这为核方法和神经网络提供了统一的泛函分析视角。近期研究从互补的视角重新激发了人们对神经网络学习动力学和泛化的泛函分析及基于范数的观点的兴趣，其中包括核极限和无限宽极限、隐式正则化以及非希尔伯特几何（Neyshabur, Bhojanapalli, & Srebro，2018年；Jacot, Gabriel, & Hongler，2018年；Sun, Gatmiry, Ahn, & Azizan，2023年）。

表 1 总结了支撑本文理论发展的结构性假设。具体而言，它阐明了为何需要特征映射诱导的再生核巴拿赫空间来表述学习问题并在巴拿赫空间中推导表示定理类结果，以及这些假设如何促成后文发展的神经网络解释。

方面	**RKHS**	**一般再生核巴拿赫空间**	**特征再生核巴拿赫空间（本文 per)**
环境函数空间	具有连续连续点求值	具有连续连续点值泛函	由.....诱导的巴拿赫空间特征映射
特征映射表示	通过再生.....而规范地存在核	并非由连续性点求值的连续性	定义中隐含
核对象	再生 核唯一确定	仅存在于.....的子类中特征再生核巴拿赫空间	与.....相关联的核特征映射
学习公式-tion	正则化 / 最小 norm 插值 via 再生核希尔伯特空间范数	可以进行正则化，但结构可能较弱	将学习表述为最小范数插值或正则化，在诱导的 RKBS
表示定理 rem	适用于范数单调正则化	取决于具体的所考虑的再生核巴拿赫空间	具有精确的结构刻画
解的形式	有限核展开于训练点	通常是非构造性的；弱-∗ 闭包问题	有限 核 展开当表示定理条件得到满足时
对偶／预对偶的作用 dual	平凡（希尔伯特自对偶）	预对偶可能不存在	预对偶 空间 is 导出的源自特征映射。在特殊的特征再生核巴拿赫空间中，预测器是一个再生核巴拿赫空间
神经网络包释	间接地，通常 via 无限宽度 or 核极限	无系统性解释	对固定架构而言自然通过诱导特征映射或相关联的核 nel

表1：再生核希尔伯特空间（再生核希尔伯特空间）、一般再生核巴拿赫空间（再生核巴拿赫空间）与本文所研究的特征再生核巴拿赫空间的比较。该表突出了在巴拿赫空间情形下获得基于核的学习表述和表示定理型解所需的结构性假设，以及它们与后文所发展的神经网络解释之间的联系。

贡献。 本文作出如下贡献:

1. 我们引入特征再生核巴拿赫空间的概念，将其与一般再生核巴拿赫空间区分开来，并刻画函数巴拿赫空间允许特征映射表示的精确结构条件。2. 我们证明，与希尔伯特空间情形不同，巴拿赫空间中点值泛函的连续性并不能保证表示定理成立，并给出再生核巴拿赫空间中表示定理型解存在的充要条件。作为推论，我们还给出了显式的充分条件。3. 我们将特征再生核巴拿赫空间中的学习表述为最小范数插值与正则化问题，并利用强调哈恩–巴拿赫定理作用的泛函分析论证，建立了解的存在性。4. 我们将该框架推广到向量值特征再生核巴拿赫空间，同时保留标量值理论的全部表达能力。因此，我们发展了相关核构造。此外，特殊特征再生核巴拿赫空间的相关性也得以明确。5. 我们证明，固定架构神经网络会自然地诱导出特殊的向量值特征再生核巴拿赫空间，从而为核方法与神经网络提供统一的泛函分析解释。最后，我们讨论训练过程，并提出代表性的训练方案。

本文聚焦于巴拿赫空间情形下学习问题的函数空间表述，以及最小范数解的存在性与结构。我们不讨论优化动力学、统计泛化界或算法效率，因为这些问题依赖于本文范围之外的其他假设。因此，我们的结果应被视为一个基础框架，可在此基础上开展此类分析。

据我们所知，这 是首篇对表示
定理在巴拿赫空间中的情形给出完整的基于特征映射的刻画，并将其系统地与固定架构神经网络联系起来的研究。

本文结构安排如下。第 2 节回顾再生核希尔伯特空间的必要背景，并将再生核巴拿赫空间置于更广泛的基于核的学习领域中加以定位。第 3 节介绍再生核巴拿赫空间的泛函分析基础，着重阐明一般再生核巴拿赫空间与由特征映射诱导的（特征）再生核巴拿赫空间之间的区别，并刻画特征表示存在所需的结构条件。第 4 节将特征再生核巴拿赫空间中的监督学习表述为最小范数插值与正则化问题，并证明解的存在性，同时建立有条件的表示定理形式。第 5节针对特殊特征再生核巴拿赫空间发展表示定理，明确保证解具有有限核表示所需的精确条件。第 6 节将该框架扩展到向量值特征再生核巴拿赫空间，并为多输出学习问题发展相关的核构造。在第 7节中，我们展示固定架构神经网络会自然诱导出特殊的向量值特征再生核巴拿赫空间，并从泛函分析角度将网络训练解释为一个最小范数学习问题。第 8 节给出一个说明性示例，将抽象构造具体化，并展示该理论如何应用于具体的神经网络架构。第9 节提炼该框架在学习理论方面的主要洞见及其影响。最后，第 10 节讨论本文的局限性、影响及未来研究方向。

2 相关工作

再生核希尔伯特空间（RKHSs）中的核方法和正则化提供了一个标准框架，在该框架下，学习可被表述为在函数空间上的优化，并且广泛的正则化经验风险最小化问题的解都可以在训练集上作有限展开。早期的表示定理类结果可追溯至样条理论（Kimeldorf & Wahba，1971年），而核学习文献中的现代形式化及推广，则在范数单调正则化和一般损失项下建立了表示定理（Schölkopf,Herbrich, & Smola，2001；Schölkopf & Smola，2002）。

为了超越希尔伯特几何——其动机来自非二次正则化、促进稀疏性的范数以及其他非希尔伯特归纳偏置——再生核巴拿赫空间（RKBSs）被引入，作为再生核希尔伯特空间的巴拿赫空间类比（Zhang 等人，2009；Xu，2023年）。基于再生核巴拿赫空间的学习面临的一个反复出现的挑战是，希尔伯特工具（例如里斯表示）通常并不可用，而表示定理需要额外的结构。近期研究继续对再生核巴拿赫空间的构造进行形式化，并阐明哪些巴拿赫几何支持基于核的学习原理和可处理的表示（Wang& Xu，2021）。
在再生核巴拿赫空间中，有限维解结构的结果可能比在再生核希尔伯特空间中复杂得多，并且可能取决于对偶结构和次微分几何。Wang 和 Xu（2021）（Wang & Xu，2021年）系统地

研究了巴拿赫空间中最小范数插值与正则化学习的表示定理，包括明确的公式，以及将优化问题化为真正有限维问题所需满足的条件。近期，Wang、Xu 和 Yan（2024年）（Wang, Xu, & Yan，2024年）研究了再生核巴拿赫空间学习的稀疏表示定理，给出了使核表示项数少于观测值数量的条件，并分析了稀疏性如何取决于再生核巴拿赫空间范数和正则化。这些研究促使我们对再生核巴拿赫空间进行更细致的分类，依据是是否存在特征映射构造和表示定理型解——这正是本文采用的视角。

与此同时，越来越多的研究通过函数空间范数和变分刻画来研究神经网络，将架构和参数范数与诱导的函数空间几何及隐式偏差联系起来（Jacot 等人，2018年；Neyshabur 等人，2018年；Wojtowytsch，2020年）。例如，神经切线核（NTK）研究提出了颇具影响力的无限宽核极限（Jacot 等人，2018年），而有关隐式正则化的研究则着重揭示，优化和几何如何在显式惩罚项之外选择解（Neyshabur 等人，2018年）。近期研究沿着与巴拿赫几何密切相关的方向，进一步搭建起核与网络之间的函数空间桥梁。Sun 等人（2023年）（Sun 等人，2023年）研究了镜像下降如何控制隐式正则化，并明确强调几何在学习得到的解中的作用。Shenouda 等人（2024年）（Shenouda、Parhi、Lee 和 Nowak，2024年）提出了向量值变差空间，这是一类专为多输出神经网络设计的再生核巴拿赫空间，并为多任务学习和压缩建立了表示定理，揭示了其结构特性。
Follain 和 Bach（2025）（Follain & Bach，2025年）提出了一种框架，通过在函数类上进行正则化经验风险最小化，明确地将特征学习与核方法结合起来。这些函数类既可以解释为核构造，也可以解释为带有学习得到的投影和非线性变换的无限宽度神经网络。
这些研究凸显了人们对统一核、神经网络与基于范数的学习原理的浓厚兴趣；在这一背景下，本文提出的特征化再生核巴拿赫空间框架提供了一种互为补充、能考虑架构的巴拿赫空间视角。

与这些研究不同，我们着重于确定在巴拿赫空间中恢复基于核的学习和表示定理所需的最小结构性假设，并探讨如何在这一框架下诠释固定架构神经网络。

3 从再生核希尔伯特空间到再生核巴拿赫空间

本节介绍再生核巴拿赫空间（再生核巴拿赫空间）的基础泛函分析框架，并重点阐述与学习理论直接相关的概念。再生核希尔伯特空间为基于核的学习提供了完备且理解充分的研究环境，但许多具有实际意义的学习问题涉及非希尔伯特几何，这些几何源于非二次正则化、诱导稀疏性的范数，或与神经网络相关的参数范数。再生核巴拿赫空间将再生核希尔伯特空间推广到巴拿赫空间，因此为研究此类情形提供了自然的数学语言，前提是满足额外的结构性假设。
我们首先回顾再生核巴拿赫空间的定义，并强调点值连续性的作用，然后介绍特征映射诱导的再生核巴拿赫空间这一概念。从学习的角度看，这一区分至关重要：与希尔伯特空间情形不同，并非每个再生核巴拿赫空间都具有特征表示，因此也并非每个再生核巴拿赫空间都支持基于核的学习算法或表示定理。本节的结果为在巴拿赫空间中表述学习问题奠定了必要的结构基础。这种表述与经典再生核希尔伯特空间理论相平行，但并非其简单延伸。

在本文中，X 将是一个非空集合，而 $\mathbb{K}$ 表示实数域 $\mathbb{R}$ 或复数域 $\mathbb{C}$。此外，$\mathcal{F}(X,\ \mathbb{K})$表示由从 X到 $\mathbb{K}$的所有函数构成的、定义在 $\mathbb{K}$ 上的向量空间，即

$$\mathcal{F}(X,\mathbb{K}) := \{f\colon X \to \mathbb{K} : f \text{ is a function}\},$$

赋予通常的逐点运算，即

$$(f+g)(x) := f(x)+g(x),\ \text{and}\ (\alpha f)(x) := \alpha f(x),$$

对所有 $f,\ g \in \mathcal{F}(X,\ \mathbb{K})$ $\alpha \in \mathbb{K}$和 $x \in X$。
本节旨在研究再生核希尔伯特空间（RKHS），为将其推广到再生核巴拿赫空间（RKBS）做准备。这两个概念分别定义在 $\mathcal{F}(X,\mathbb{K})$ 的希尔伯特空间或巴拿赫空间上，它们是向量子空间，分别记为 H 和 B，因此其中的元素都是函数，且不预设任何性质。这些空间的基域为 $\mathbb{K}$。因此，$H \subseteq \mathcal{F}(X,\mathbb{K})$ 表示希尔伯特空间 H 是 $\mathcal{F}(X,\mathbb{K})$ 的向量子空间；巴拿赫空间 $B \subseteq \mathcal{F}(X,\mathbb{K})$ 的情况也类似。本文将始终采用这一约定，不再另行说明（例如，参见定义 3.2）。在下一个定理中，我们将刻画一个再生

核希尔伯特空间（RKHS）的特征映射和核函数。在此语境下，若 $\mathbb{K}=\mathbb{R}$，则共轭视为恒等变换。

定理 3.1。 设 $H \subseteq \mathcal{F}(X,\mathbb{K})$ 是一个希尔伯特空间。那么，以下断言等价：

(i) 对于每个 $f \in H$，由 $\delta_x(f)=f(x)$ 给出的线性映射 $\delta_x\colon H \to \mathbb{K}$ 对每个 $x \in X$ 都是连续的。这意味着存在 $C_x \geq 0$，使得

$$|\delta_x(f)| = |f(x)| \leq C_x \|f\|, \text{ for every } f \in H.$$

(ii) 存在一个映射 $\Phi\colon X \to H$，使得 $f(x)=\langle f, \Phi(x)\rangle$，对每个 $x \in X$.

(iii) 存在一个函数 $k\colon X \times X \to \mathbb{K}$，使得 $k(\cdot,x) \in H$，对每个 $x \in X$，满足 $f(x)=\langle f, k(\cdot,x)\rangle$，对每个 $f \in H$ 和 $x \in X$.

如果这些断言成立，则 Φ 与 k 是唯一的。此外，

$$H = \overline{\mathrm{Lin}\{k(\cdot,x) : x \in X\}}^{\|\cdot\|} = \overline{\mathrm{Lin}\{\Phi(x) : x \in X\}}^{\|\cdot\|}$$

and

$$k(x,x') = \overline{\langle k(\cdot,x), k(\cdot,x')\rangle} = \overline{\langle \Phi(x), \Phi(x')\rangle}.$$

证明。 (i) $\Longrightarrow$ (ii) 由于 δ_x 是线性的，其连续性蕴含 $\delta_x \in H^*$ 从而，根据里斯表示定理（Conway，1994，定理 3.4），存在唯一的 $k_x \in H$ 使得 $\delta_x(\cdot)=\langle \cdot,\ k_x\rangle$. 令 $\Phi\colon X \to H$由下式给出 $\Phi(x)=k_x$，对每一个 $x \in X$.则，

$$\langle f, \Phi(x)\rangle = \langle f, k_x\rangle = \delta_x(f) = f(x), \text{ where } f \in H.$$

(ii) $\Longrightarrow$ (iii) 定义 $k(\cdot,x) := \Phi(x)$以及 $k(x,x') = \delta_x(k(\cdot,x')) = \delta_x(\Phi(x'))$，对每一个 $x,x' \in X$（注意$k(\cdot,x') = \Phi(x') \in H \subseteq \mathcal{F}(X,\mathbb{K})$）。由(ii)，我们有 $k(\cdot,x) \in H$ 以及 $f(x)=\langle f, k(\cdot,x)\rangle$，对每一个 $x \in X$.

(iii) $\Longrightarrow$ (i) 由于 $f(x)=\langle f,\ k(\cdot,\ x)\rangle$，对于每个 $x \in X$ 和 $f \in H$，由柯西-施瓦茨不等式，令$C_x = \|k(\cdot,\ x)\|$，可得

$$|\delta_x(f)| = |f(x)| \leq C_x \|f\|,$$

对于每个 $x \in X$，这就证明了 (i)。

最后，$\Phi(x)$ 的唯一性源于以下事实：$\langle f,\ \Phi(x)\rangle = f(x)$。事实上，如果另一个映射 $\Phi\colon X \to H$ 也满足这一性质，那么

$$\left\langle f, \Phi(x) - \widetilde{\Phi}(x)\right\rangle = f(x) - f(x) = 0,$$

对于每个 $f \in H$ 和 $x \in X$. 取 $f = \Phi(x) - \widetilde{\Phi}(x) \in H$，我们得到 $\Phi = \widetilde{\Phi}$。

k 的唯一性证明类似，只需分别用 $k(\cdot,\ x)$ 和 $k(\cdot,\ x)$ 替换 $\Phi(x)$ 和 $\widetilde{\Phi}(x)$，最后，由于

$$f(x) = \langle f, \Phi(x)\rangle = \langle f, k(\cdot,x)\rangle,$$

对每个 $x \in X$ 和 $f \in H$，可知，若 $\langle f,\ \Phi(x)\rangle = 0$ 对每个 $x \in X$ 则 $(\mathrm{Lin}\{\Phi(x) :\ x \in X\})^{\perp} = \{0\}$ 因此，根据（Conway，1994，推论 2.9），可知

$$H = \{0\}^{\perp} = (\mathrm{Lin}\{\Phi(x) : x \in X\})^{\perp\perp} = \overline{\mathrm{Lin}\{\Phi(x) : x \in X\}}^{\|\cdot\|}.$$

按照相同的推理，$\{k(\cdot,\ x) :\ x \in X\}^{\perp} = \{0\}$ 因此，

$$H = \overline{\mathrm{Lin}\{k(\cdot,x) : x \in X\}}^{\|\cdot\|}.$$

最后，

$$k(x,x') = \delta_x(k(\cdot,x')) = \langle k(\cdot,x'), k(\cdot,x)\rangle = \overline{\langle k(\cdot,x), k(\cdot,x')\rangle},$$

类似地，$k(x,\ x') = \delta_x(\Phi(x')) = \langle \Phi(x')\ \Phi(x)\rangle = \langle \Phi(x)\ \Phi(x')\rangle$. □

下一个定义可见于（Paulsen 与 Raghupathi，2016，定义 2.1）。

定义 3.2。再生核希尔伯特空间（**再生核希尔伯特空间**）是一个希尔伯特空间 $H \subseteq \mathcal{F}(X, \mathbb{K})$，其点求值映射 $\delta_x : H \to \mathbb{K}$ 定义为

$$\delta_x(f) = f(x), \quad \text{for every} \quad f \in H,$$

对每个 $x \in X$.都是连续的

如果 H 是一个再生核希尔伯特空间，那么定理 3.1 的断言 (ii) 中给出的映射 Φ 称为与 H相关联的特征映射；类似地，断言 (iii) 中给出的映射 k 称为与 H.相关联的再生核函数。此外，核 k的再生性质就是等式 $f(x) = \langle f, k(\cdot, x)\rangle$，对任意 $x \in X$和 $f \in H$.均成立

显然，在定义 3.2 中， δ_x 的连续性（即定理 (i)中的断言 3.1）可以由定理 (ii) 或 (iii)中的断言 3.1替代，因为正如我们所证明的，这三个性质是等价的。

备注 3.3。摩尔—阿龙尚定理（Aronszajn，1950，性质 2.4）指出，对于每个厄米正定函数 k : $X \times X \to \mathbb{K}$，都存在唯一的再生核希尔伯特空间（RKHS）$H \subseteq \mathcal{F}$（$X$，$\mathbb{K}$），反之亦然。这一对应关系源于以下事实：当且仅当空间 k 是厄米正定的

$$H_0 := \text{Lin}\{k(x, \cdot) : x \in X\}$$

是一个预希尔伯特空间。事实上，如果 H 是一个再生核希尔伯特空间，那么相应的再生核必然是厄米且正定的，这一点可以直接验证。反过来，如果 k： $X \times X \to \mathbb{K}$ 是厄米且正定的，那么由 H 补全 $H_0 :=$ $\text{Lin}\{k(x, \cdot) \text{ 构成的 } x \in X\}$ 空间是一个希尔伯特空间，因此 k 满足定理(iii)中的条件3.1。

上述 定义引出了下面的定义，该定义见（Lin、Zhang 和 Zhang，2022，定义 2.1）。

定义 3.4。巴拿赫空间 $B \subseteq \mathcal{F}(X, \mathbb{K})$ 称为**再生核巴拿赫空间**（**再生核巴拿赫空间**），如果对每个 $x \in X$，求值映射 $\delta_x : B \to \mathbb{K}$ 定义为

$$\delta_x(f) = f(x) \quad (f \in B),$$

是连续的。也就是说，存在 $C_x \geq 0$ 使得

$$|\delta_x(f)| = |f(x)| \leq C_x \|f\|, \text{ for every } f \in B.$$

显然，每个再生核希尔伯特空间都是再生核巴拿赫空间。其他重要的例子也很容易得到，如下所示。

例 3.5。设 X 是一个非空集合， W 是一个巴拿赫空间。那么，每个映射 $\Phi : X \to W$ 都可以构造出一个再生核巴拿赫空间。事实上，若

$$W_0 := \overline{\text{Lin}\{\Phi(x) : x \in X\}}^{\|\cdot\|},$$

则由 B_Φ 给出的空间为

$$B_\Phi := \{f_\mu(\cdot) := \mu(\Phi(\cdot)) : \mu \in W_0^*\} \subseteq \mathcal{F}(X, \mathbb{K}),$$

其中 $f_\mu(x) := \mu(\Phi(x))$，赋予范数 $\|f_\mu\| := \|\mu\|$，后，是一个再生核巴拿赫空间。

B_Φ 是一个向量空间，这一点显然，因为对于每个 μ， $\tilde{\mu} \in W_0^*$和 $\alpha \in \mathbb{K}$.， $f_\mu + f_{\tilde{\mu}} = f_{\mu+\tilde{\mu}}$且 $\alpha f_\mu = f_{\alpha\mu}$。由于 B_Φ在代数上同构于 W_0^*，因此显然 $\|f_\mu\| := \|\mu\|$定义了 B_Φ上的巴拿赫范数。另一方面，求值映射 $\delta_x : B_\Phi \to \mathbb{K}$满足

$$|\delta_x(f_\mu)| = |\mu(\Phi(x))| \leq \|\Phi(x)\| \, \|\mu\|.$$

因此， $\delta_x(f_\mu)| \leq C_x \|f_\mu\|$，对于每个 $f_\mu \in B_\Phi$，其中 $C_x = \|\Phi(x)\|$。因此， B_Φ 是一个再生核巴拿赫空间 .

我们只关注那些 B_Φ 与映射 Φ： $X \to W$相关联的对象，其中 W 是一个巴拿赫空间，使得$W \subseteq \mathcal{F}(X', \mathbb{K})$对某个非空集合 X'成立。这引出了下面的定义。

定义 3.6。特征映射 **是一个映射** Φ： $X \to W \subseteq \mathcal{F}$（$X'$，$\mathbb{K}$），其中 X 和 X' 是非空集合，并且 W 是 $\mathcal{F}$（X'，$\mathbb{K}$）的一个向量子空间，同时自身也是一个巴拿赫空间。如果 Φ 是一个特征映射，那么例 B_Φ 3.5 中描述的空间称为**与 Φ.相关联的再生核巴拿赫空间**

正如我们将要展示的，与特征映射这一概念相关联的是核函数的概念。

定义 3.7。设 X 和 X' 为非空集合。如果存在一个巴拿赫空间 $W \subseteq \mathcal{F}$（X'，$\mathbb{K}$），使得对于每个 $x \in X.$， k（x，$\cdot$） $\in W$，，则称映射 $k : X \times X' \to \mathbb{K}$ 为一个**核函数**。此时，将空间 W 称为与 k 相关联的**巴拿赫空间**。

命题 3.8。设 X 和 X' 为非空集合，且 $W \subseteq \mathcal{F}(X', \mathbb{K})$ 为一个巴拿赫空间。那么，每个满足对于任意 $x \in X$ 都有 $k: X \times X' \to \mathbb{K}$，$k(x, \cdot) \in W$ 的核函数，都有唯一的相关联特征映射 $\Phi: X \to W \subseteq \mathcal{F}(X', \mathbb{K})$，反之亦然。

证明。给定核函数 $k: X \times X' \to \mathbb{K}$，若对于每个 $x \in X$，都有 $k(x, \cdot) \in W \subseteq \mathcal{F}(X', \mathbb{K})$，则其相关特征映射定义为 $\Phi: X \to W$ 其中，对于每个 $x \in X$.，$\Phi(x) := k(x, \cdot)$
反过来，给定特征映射 $\Phi: X \to W \subseteq \mathcal{F}(X', \mathbb{K})$，将其相关核函数定义为 $k: X \times X' \to \mathbb{K}$，由 $k(x, x') = \delta'_x(\Phi(x))$ 给出，其中对于每个 $x \in X$ 和 $x' \in X'$，$k(x, \cdot) := \Phi(x)$。 □

注 3.9。根据上述命题，若 B_Φ 是与特征映射 $\Phi: X \to W$ 相关联的再生核巴拿赫空间，那么，B_Φ 也可以从核的角度来定义。也就是说

$$B_\Phi^k := \{\widehat{f_\mu}(\cdot) : \mu \in W_0^*\},$$

其中 $\widehat{f_\mu}(x) := \mu(k(x, \cdot))$ 且 $\left\|\widehat{f_\mu}\right\| := \|\mu\|$。因此，$B_\Phi^k = B_\Phi$ 其中 $\Phi(x) := k(x, \cdot)$，对于每个 $x \in X$，我们称 B_Φ^k 为**与核函数相关联的再生核巴拿赫空间** k.

我们的下一个目标是证明，与再生核希尔伯特空间的情形不同，并非所有再生核巴拿赫空间都与特征映射相关 .

从现在起，给定一个巴拿赫空间 W，我们将用 W^*表示其拓扑对偶，并用 $J: W \to W^{**}$ 表示由下式定义的典范嵌入 $J(w) = J_w$，其中 $J_w(f) = f(w)$ 对每个 $f \in W^*$. 众所周知，J是等距映射；也就是说，$\|J_w\| = \|w\|$ 对每个 $w \in W$ 其中 W^{**} 上的范数是标准算子范数（例如，参见（Megginson，1998，命题 1.11.3）。

如（Dales 与 Ülger，2025年，定义 1.3.24）所述：

定义 3.10。设 B 是一个巴拿赫空间。B 的一个**预对偶**是一个有序对 (B_*, τ)，由一个巴拿赫空间 B_* 和一个线性同胚 $\tau: B \to (B_*)^*$. 组成。如果 τ 是等距映射，则称该预对偶是等距的。如果空间 B 存在一个预对偶 (B_*, τ)，则称其为**对偶**空间。

本文中我们只考虑等距前对偶 (B_*, τ)，这意味着映射 $\tau: B \to (B_*)^*$ 是一个等距同构。受此启发，我们引入以下定义。

定义 3.11。设 W_1 和 W_2 是巴拿赫空间。如果存在一个等距同构 $\varphi: W_1 \to W_2$，则称 W_1 是 W_2 的**副本**，此时记作 $W_1 \overset{\varphi}{\equiv} W_2$。因此，$\varphi$是一个线性双射映射，满足$\|\varphi(w_1)\| = \|w_1\|$，对每个 $w_1 \in W_1$ 都成立。此外，如果 $w_1 \in W_1$ 和 $w_2 \in W_2$ 满足 $\varphi(w_1) = w_2$，则记作$\varphi w_1 \equiv w_2$。

很多时候，当等距同构显而易见（或无关紧要）时，我们就简单地写作 $W_1 \equiv W_2$ 和 $w_1 \equiv w_2$。
回顾一下，若 W_1 和 W_2 是巴拿赫空间，且 $\varphi: W_1 \to W_2$ 是一个同构，那么 φ的伴随映射，即由下式定义的映射 $\varphi^*: W_2^* \to W_1^*$

$$[\varphi^*(w_2^*)](w_1) := w_2^*(\varphi(w_1)),$$

对每个 $w_1 \in W_1$，它也是一个同构（详见（vanNeerven，2022，命题 4.27））。因此，我们得到以下结果，其证明是显然的。

命题 3.12。设 W_1和 W_2为巴拿赫空间。若 W_1是 W_2的一个副本，则 W_2^*是 W_1^*的一个副本。事实上，若$W_1 \overset{\varphi}{\equiv} W_2$，则 $W_2^* \overset{\varphi^*}{\equiv} W_1^*$。此外，若 $w_i^* \in W_i^*$对 $i = 1$成立，2，则 $w_2^* \equiv w_1^*$当且仅当 $w_1^* = w_2^*\varphi$。

注意，如果 $B \subseteq \mathcal{F}(X, \mathbb{K})$ 是一个具有等距前对偶 B_* 的再生核巴拿赫空间 $B \overset{\varphi}{\equiv} (B_*)^*$，并且根据命题 3.12，我们还有 $(B_*)^{**} \overset{\varphi^*}{\equiv} B^*$。

接下来的结果刻画了再生核巴拿赫空间中特征映射的存在性。

命题 3.13。设 $B \subseteq \mathcal{F}(X, \mathbb{K})$ 是一个再生核巴拿赫空间，且设 B_* 是一个巴拿赫空间，使得 $B \overset{\varphi}{\equiv} (B_*)^*$。则以下性质等价：

(i) 对于每个 $x \in X$，都有 $J_{w_x} \overset{\varphi^*}{\equiv} \delta_x$，其中某个 $w_x \in B_*$.

(ii) 存在一个映射 $\Phi: X \to B_*$，使得对于每个 $x \in X$，均有 $J_{\Phi(x)} \overset{\varphi^*}{\equiv} \delta_x$。

此外，如果满足这些性质，那么 $B_* = \mathrm{Lin}\{\Phi(x)\colon\ x \in X\}^{\|\cdot\|}$ 并且

$$B = B_\Phi := \{f_\mu(\cdot) := \mu(\Phi(\cdot)) : \mu \in (B_*)^*\}.$$

证明。 (ii) $\Longrightarrow$ (i) 这一陈述显然成立（对每个 $x \in X$，令 $w_x := \Phi(x)$，）。

(i) $\Longrightarrow$(ii) 首先，我们注意到，若对于某个 $w_x \in B_*$，$J_{w_x} \overset{\varphi^*}{\equiv} \delta_x$，则这样的元素 w_x 必定唯一。事实上，若存在 w_x，$\widetilde{w}_x \in B_*$ 使得 $J_{w_x} \overset{\varphi^*}{\equiv} \delta_x$ 且 $J_{\widetilde{w}_x} \overset{\varphi^*}{\equiv} \delta_x$，那么 $J_{w_x} = J_{\widetilde{w}_x} = (\varphi^*)^{-1}(\delta_x)$，因此 $J_{w_x - \widetilde{w}_x} = 0$，这意味着 $w_x = \widetilde{w}_x$。因此，我们可以将 $\Phi\colon X \to B_*$ 定义为 $\Phi(x) = w_x$，对每个 $x \in X$.

最后， $B = B_\Phi$ 这是由以下事实得出的： $J_{\Phi(x)} \overset{\varphi^*}{\equiv} \delta_x$。根据命题3.12 这意味着 $J_{\Phi(x)}\varphi = \delta_x$。我们声称，空间$\mathrm{Lin}\{\Phi(x)\colon\ x \in X\} \subseteq B_*$ 满足

$$(\mathrm{Lin}\{\Phi(x) : x \in X\})^\perp = \{0\}.$$

事实上，由于 $(B_*)^* = \varphi(B)$，我们推得，若 $f \in B$ 满足 $\varphi(f)(\Phi(x)) = 0$ 对每个 $x \in X$均成立，则$J_{\Phi(x)}(\varphi(f)) = 0$对每个 $x \in X$均成立，因此 $\delta_x(f) = 0$对每个 $x \in X$均成立，进而 $f = 0$，从而 $\varphi(f) = 0$，这就证明了该论断。根据（Conway，1994，推论 2.9），我们有

$$B_* = \overline{\mathrm{Lin}\{\Phi(x) : x \in X\}}^{\|\cdot\|}.$$

因此，对每个 $f \in B$,

$$f_{\varphi(f)}(x) := \varphi(f)(\Phi(x)) = J_{\Phi(x)}(\varphi(f)) = \delta_x(f) = f(x) \qquad (x \in X),$$

因此 $f = f_{\varphi(f)}$，进而

$$B := \{f : f \in B\} = \{f_{\varphi(f)} : f \in B\} = \{f_{\varphi(f)} : \varphi(f) \in (B_*)^*\} = \{f_\mu : \mu \in (B_*)^*\} = B_\Phi,$$

如所欲证。 □

我们已经证明，若 $B \subseteq \mathcal{F}(X,\ \mathbb{K})$ 是再生核巴拿赫空间，则对每一个 $x \in X$ 都有 $\delta_x \in B^*$，且存在两种可能：要么 B 容许一个预对偶 B_*，使得 $B \equiv (B_*)^*$，并且 δ_x 可与 $J(B_*) \subseteq (B_*)^{**} \equiv B^*$，中的唯一元素等同；要么不存在这样的预对偶。

第一种可能意味着存在 $\Phi\colon\ X \to B_*$，且 $B_* := \mathrm{Lin}\{\Phi(x)\colon\ x \in X\}^{\|\cdot\|}$，使得 $B = B_\Phi$.。这为如下定义提供了依据。

定义 3.14. 设 $B \subseteq \mathcal{F}(X,\ \mathbb{K})$ 为再生核巴拿赫空间。若对某个特征映射 $\Phi\colon\ X \to W \subseteq \mathcal{F}(X',\ \mathbb{K})$ 有 $B = B_\Phi$，则称 B 为**特征再生核巴拿赫空间**。这意味着

$$B = B_\Phi := \{f_\mu(\cdot) := \mu(\Phi(\cdot)) : \mu \in W_0^*\},$$

其中 $W_0 := \overline{\mathrm{Lin}\{\Phi(x)\colon\ x \in X\}}^{\|\cdot\|}$，且 $\|f_\mu\| := \|\mu\|$.

注意， $B = B_\Phi \overset{\varphi}{\equiv} W_0^*$ 其中 $\varphi(f_\mu) = \mu$ （参见例 3.5）。

以下结果由命题 3.13推得。

推论 3.15。 设 $B \subseteq \mathcal{F}(X, \mathbb{K})$ 是一个再生核巴拿赫空间，并且对每个 $x \in X$，令 $\delta_x\colon\ B \to \mathbb{K}$ 为由 $\delta_x(f) := f(x)$给出的求值映射，对所有 $f \in B$。那么， B是一个特征再生核巴拿赫空间，当且仅当 B有一个预对偶 B_*，使得 $B \overset{\varphi}{\equiv} (B_*)^*$，并且对每个 $x \in X$，存在 $w_x \in B_*$ 使得 $J_{w_x} \overset{\varphi^*}{\equiv} \delta_x$ （即 $\delta_x = J_{w_x}\varphi$），其中 $J\colon\ B_* \to (B_*)^{**}$ 是典范嵌入。

*Pro*的。 若 $B = B_\Phi$ 对于某个特征映射 $\Phi\colon X \to W \subseteq \mathcal{F}(X', \mathbb{K})$，则 $B = B_\Phi \overset{\varphi}{\equiv} W_0^*$，其中

$$W_0 := \overline{\mathrm{Lin}\{\Phi(x) : x \in X\}}^{\|\cdot\|},$$

且对于每个 $f_\mu \in B_\Phi$， $\varphi(f_\mu) = \mu$。此外，对于每个 $\mu \in W_0^*$， $f_\mu \in B_\Phi$，以及 $x \in X$，

$$\delta_x(f_\mu) = f_\mu(x) = \mu(\Phi(x)) = J_{\Phi(x)}(\mu) = J_{\Phi(x)}(\varphi(f_\mu)),$$

使得 $\delta_x = J_{\Phi(x)}\varphi$ 这意味着，根据命题 3.12， $J_{\Phi(x)} \overset{\varphi^*}{\equiv} \delta_x$对每个 $x \in X$。最后，令 $B_* := W_0$且 $w_x := \Phi(x)$从而完成必要条件的证明。充分条件直接由命题 3.13得出。

□

下一个结果总结了我们讨论过的所有等价事实。

定理 3.16。 设 $B \subseteq \mathcal{F}(X, \mathbb{K})$ 是一个巴拿赫空间。考虑以下断言：

(i) B 是一个再生核巴拿赫空间（即，对于每个 $x \in X$，点求值映射 $\delta_x \colon B \to \mathbb{K}$，定义为 $\delta_x(f) = f(x)$，对$f \in B$而言是连续的；因此 $\delta_x \in B^*$）。

(ii) B 是一个特征再生核巴拿赫空间（即，存在一个特征映射 $\Phi \colon X \to W \subseteq \mathcal{F}(X', \mathbb{K})$，使得 $B = B_\Phi$）。

(iii) B 有一个预对偶 B_*，使得 $B \overset{\varphi}{=} (B_*)^*$，并且对于每个 $x \in X$，存在 $w_x \in B_*$，使得 $J_{w_x} \overset{\varphi^*}{=} \delta_x$。

(iv) B 有一个预对偶 B_*，使得 $B \overset{\varphi}{=} (B_*)^*$，并且存在一个特征映射 $\Phi \colon X \to B_*$，使得$B_* = \operatorname{Lin}\{\Phi(x) : \overline{x \in X\}}^{\|\cdot\|}$ 且 $J_{\Phi(x)} \overset{\varphi^*}{=} \delta_x$，对每个 $x \in X$.

(v) B 有一个预对偶 B_*，使得 $B \overset{\varphi}{=} (B_*)^*$，并且存在一个核函数 $k \colon X \times X' \to \mathbb{K}$，使得 $B_* := \operatorname{Lin}\{k(x, \cdot) : \overline{x \in X\}}^{\|\cdot\|}$ 且 $J_{k(x,\cdot)} \overset{\varphi^*}{=} \delta_x$，对每个 $x \in X$。

(vi) 存在一个核函数 $k \colon X \times X' \to \mathbb{K}$，使得 $B = B_\Phi^k$.

于是(ii) $\iff$ (iii) $\iff$ (iv) $\iff$ (v) $\iff$ (vi) $\implies$ (i)；若 B_* 是自反的，则这些断言全部等价。此外，若上述断言成立，则 Φ和 k 是唯一的。

证明。 由命题 3.13和推论3.15 可知(ii) $\iff$ (iii) $\iff$ (iv)。此外，等价关系(iv) $\iff$ (v) 由命题3.8推出，而 (ii) $\iff$ (vi) 则由注释 3.9得出。因此，

$$\text{(ii)} \iff \text{(iii)} \iff \text{(iv)} \iff \text{(v)} \iff \text{(vi)}.$$

显然，(ii) $\Rightarrow$ (i)。 此外，若 B_* 是自反的，则 $B_* \overset{J}{=} (B_*)^{**}$，因此断言 (i) $\Rightarrow$ (iii) 显然成立，因为$\delta_x \in B^* \equiv (B_*)^{**} = J(B_*)$。

最后，若 $B \overset{\varphi}{=} (B_*)^*$ 且 Φ，$\widetilde{\Phi} \colon X \to B_*$ 满足这样的条件

$$B_* = \overline{\operatorname{Lin}\{\Phi(x) : x \in X\}}^{\|\cdot\|} = \overline{\operatorname{Lin}\{\widetilde{\Phi}(x) : x \in X\}}^{\|\cdot\|}$$

且 $J_{\Phi(x)} \overset{\varphi^*}{=} \delta_x \overset{\varphi^*}{=} J_{\widetilde{\Phi}(x)}$，对每个 $x \in X$，都有 $J_{\Phi(x)} - J_{\widetilde{\Phi}(x)} = 0$，因此 $\Phi = \widetilde{\Phi}$. 同样，k 也是唯一的。 □

我们已在定义 3.14 中确立，巴拿赫空间 $B \subseteq \mathcal{F}(X, \mathbb{K})$ 是特征再生核巴拿赫空间，当且仅当 $B = B_\Phi$存在某个特征映射 $\Phi : X \to W \subseteq \mathcal{F}(X, \mathbb{K}')$。 在这种情况下，我们称 Φ 为与之关联的**特征映射**$B = B_\Phi$，而相应的核 $k \colon X \times X' \to \mathbb{K}$，满足 $B = B_\Phi^k = B_\Phi$，是**核**

与 $B = B_\Phi$相关联的函数。此外，性质 $J_{k(x,\cdot)} \overset{\varphi}{=}{}^* \delta_x$（等价地 $J_{\Phi(x)}$）对每个 $x \in X$ 都成立，称为**再生性质**。

注 3.17。 如前所述， $B \subseteq \mathcal{F}(X, \mathbb{K})$ 是特征再生核巴拿赫空间，当且仅当 $B = B_\Phi$ 存在某个特征映射$\Phi \colon X \to W$，在这种情况下

$$W_0 = \overline{\operatorname{Lin}\{\Phi(x) : x \in X\}}^{\|\cdot\|} = B_*.$$

在许多应用中，巴拿赫空间 W 可能是已知的，而预对偶 $W_0 = B_*$ 却并不为人所知。受这一观察的启发，我们还通过空间 W^*. 来定义 B_Φ。具体如下：

$$B_\Phi := \{f_\mu(\cdot) := \mu(\Phi(\cdot)) : \mu \in W^*\}$$

其中 $\|f_\mu\| := \inf\{\|\widetilde{\mu}\| : \widetilde{\mu} \in W^* \text{ 满足 } f_\mu = f_{\widetilde{\mu}}\}$. 事实上，

$$\|f_\mu\| := \inf\{\|\widetilde{\mu}\| : \widetilde{\mu} \in W^* \text{ with } f_\mu = f_{\widetilde{\mu}}\} = \left\|\mu + W_0^\perp\right\| = \left\|\mu_{|W_0}\right\|.$$

注意， $W^*/W_0^\perp \equiv W_0^*$根据经典的同一对应 $\mu + W_0^\perp \equiv \mu_{|W_0}$（例如，参见（Megginson，1998，定理 1.10.16））。

众所周知，巴拿赫空间 $B \subseteq \mathcal{F}$（X，$\mathbb{K}$）是再生核巴拿赫空间，当且仅当 $\delta_x \in B^*$，对每个 $x \in X$ 并且我们已经证明，若 B 没有预对偶，则 B 不是特征再生核巴拿赫空间。因此，存在并非特征再生核巴拿赫空间的再生核巴拿赫空间，这与再生核希尔伯特空间的情形相反（参见定理（i）和（ii）中的等价关系3.1）。

为了说明这一点，我们给出以下例子。

例 3.18。 设 $B = C_0(\Omega)$，其中 Ω 是一个非紧的局部紧豪斯多夫空间。那么， B 没有预对偶，但求值映射 δ_x 仍然是连续的，对每个 $x \in X$.均如此。因此， $\delta_x \in B^*$， 对每个$x \in X$， 都成立，并且不存在一个巴拿赫空间 W 及特征映射 $\Phi\colon X \to W$ 使得 $B = B_\Phi$. 这是因为，若情况如此，则 $B = C_0(\Omega)$与之等距同构

$$(B_*)^* = \left(\overline{\mathrm{Lin}\{\Phi(x) : x \in X\}}\right)^*.$$

因此，由于巴拿赫空间之间的等距同构会保持闭单位球的极点，可知 $C_0(\Omega)$ 的闭单位球有极点，这就产生了矛盾。事实上，根据巴拿赫–阿劳格鲁定理（Rudin，1991，定理 3.15），闭单位球 $B(0,1)$ 在 $(B_*)^*$ 是弱*–紧的；根据克莱因–米尔曼定理（Rudin，1991，定理 3.23）， $B(0,1)$ 的极点集非空，且 $B(0,1)$ 是它们的弱*–闭凸包。

现在证明 $C_0(\Omega)$ 的闭单位球没有极点。考虑 $f \in C_0(\Omega)$，其中 $\|f\|_\infty = 1$ (若$\|f\|_\infty < 1$，那么 f显然不是极点)。令 $F_1 := |f|^{-1}([0, \frac{1}{4}])$，且 $F_2 := |f|^{-1}([\frac{3}{4}, 1])$。由于 $|f|$ 连续， F_1 和 F_2是 Ω 的闭子集。此外， $F_1 \neq \varnothing$，因为存在一个紧集 $K \subseteq \Omega$，使得对于每个 $x \in \Omega\backslash K.$，都有 $|f(x)| < \frac{1}{4}$。又因为 $\|f\|_\infty = 1.$，所以 $K \neq \varnothing$。类似地，由于 $\|f\|_\infty = 1.$，所以 $F_2 \neq \varnothing$。显然，$F_1 \cap F_2 = \varnothing$，而乌雷松引理（Naimark，1972，第 28 页）允许我们找到 $g : \Omega \to [0, \frac{1}{4}]$，使得对于每个 $x \in F_1$，都有 $g(x) = \frac{1}{4}$；对于每个 $x \in F_2.$，都有 $g(x) = 0$。由此可得，对于每个 $x \in \Omega$， $|(f \pm g)(x)| \leq 1$。因此，$\|f \pm g\|_\infty = 1$，并且

$$f = \frac{1}{2}(f + g) + \frac{1}{2}(f - g),$$

这证明了 f 不是 $C_0(\Omega)$ 单位球的极端点。由此可得，对于任意巴拿赫空间 $B_*.$， $C_0(\Omega)$ 都不与 $(B_*)^*$ 等距同构。

注释 3.19。巴拿赫空间的预对偶不一定唯一。例如， c 和 c_0 是 l_1 的非同构预对偶（参见 Brown & Ito, 1976; Rao, 2009）。从这个意义上说， $B \subseteq \mathcal{F}(X, \mathbb{K})$可以是一个特征巴拿赫空间，对应于与不同预对偶相关的$\widetilde{}$个不同特征映射。然而，如果 B_* 和 B_* 是 B 的预对偶（它们不一定$\widetilde{}$同构），且满足 $B \overset{\varphi}{\equiv} (B_*)^*$ 和 $B \overset{\widetilde{\varphi}}{\equiv} (\widetilde{B}_*)^*$，并且 $\Phi\colon X \to B_*$ 和 $\widetilde{\Phi}\colon X \to \widetilde{B_*}$ 满足

$$B_* := \overline{\mathrm{Lin}\{\Phi(x) : x \in X\}}^{\|\cdot\|} \text{ and } \widetilde{B_*} := \overline{\mathrm{Lin}\{\widetilde{\Phi}(x) : x \in X\}}^{\|\cdot\|}$$

$\widetilde{}$那么， $B_\Phi \equiv (B_*)^*$ 和 $B_{\widetilde{\Phi}} \equiv (B_*)^*$，使得 $B_\Phi \equiv B \equiv B_{\widetilde{\Phi}}$。

此外，如果 $\delta_x \overset{\varphi^*}{\equiv} J_{\Phi(x)}$ 且 $\delta_x \overset{\widetilde{\varphi}^*}{\equiv} J_{\widetilde{\Phi}(x)}$ （这意味着 $\delta_x = J_{\Phi(x)}\varphi = J_{\widetilde{\Phi}(x)}\widetilde{\varphi}$），对每个 $x \in X$，则 $B = B_\Phi = B_{\widetilde{\Phi}}$。

我们知道，如果一个巴拿赫空间 $B \subseteq \mathcal{F}(X, \mathbb{K})$ 是一个特征再生核巴拿赫空间，那么 B 具有一个预对偶 B_*，并且对于某个特征映射 $\Phi\colon X \to B_*$，有 $B = B_\Phi$，使得

$$B_* := \overline{\mathrm{Lin}\{\Phi(x) : x \in X\}}^{\|\cdot\|}$$

（见定理3.16）。此外，若 B_* 是一个再生核巴拿赫空间，则可得到以下结果。

定理 3.20。 设 $B_\Phi \subseteq \mathcal{F}(X, \mathbb{K})$ 是一个特征再生核巴拿赫空间，其预对偶为 $B_* \subseteq \mathcal{F}(X', \mathbb{K})$。设 $\Phi\colon X \to B_*$ 为相应的特征映射， $k\colon X \times X' \to \mathbb{K}$ 为相关核。若 B_* 是一个再生核巴拿赫空间，则对每个 $x' \in X'$，都有 $k(\cdot, x') \in B_\Phi$ 和 $\delta_{x'} \in (B_*)^*$，其中 $f_{\delta_{x'}} = k(\cdot, x')$。此外，

$$B_\Phi = \overline{\mathrm{Lin}\{k(\cdot, x') : x' \in X'\}}^{w*} \equiv \overline{\mathrm{Lin}\{\delta_{x'} : x' \in X'\}}^{w*} = (B_*)^*,$$

其中 $k(\cdot, x') \equiv \delta_{x'}$。

证明。 B_* 是一个再生核巴拿赫空间，意味着对每个 $x' \in X'$， 都有 $\delta_{x'} \in (B_*)^*$，其中

$$B_* := W_0 := \overline{\mathrm{Lin}\{\Phi(x) : x \in X\}}^{\|\cdot\|}.$$

根据定义 3.14，我们有 $B_\Phi = \{f_{\mu_0}(\cdot) := \mu_0(\Phi(\cdot)) : \ \mu_0 \in (B_*)^* = (W_0)^*\}$，其中 $\|f_{\mu_0}\| = \|\mu_0\|$. 在由以下方式给出的 $)^* = W_0^* \equiv B_\Phi$ 识别下

$$\mu_0 \in W_0^* \equiv f_{\mu_0}(\cdot) \in B_\Phi,$$

我们有 $\delta_{x'} \in W_0^* \equiv f_{\delta_{x'}}(\cdot) = k(\cdot, x') \in B_\Phi$，因此

$$\mathrm{Lin}\{\delta_{x'} : x' \in X'\} \subseteq (B_*)^* = W_0^* \equiv \mathrm{Lin}\{k(\cdot, x') : x' \in X'\} \subseteq B_\Phi,$$

并且我们声称

$$\overline{\mathrm{Lin}\{k(\cdot,x') : x' \in X'\}}^{w*} = B_\Phi \equiv \overline{\mathrm{Lin}\{\delta_{x'} : x' \in X'\}}^{w*} = (B_*)^*. \tag{3.1}$$

为证明该论断，注意到这个识别$\delta_{x'} \equiv f_{\delta_{x'}}(\cdot)$显然。此外，$f_{\delta_{x'}}(x) = \delta_{x'}(\Phi(x))$并且，由于$\Phi(x) = k(x,\cdot)$根据命题3.8，我们有

$$f_{\delta_{x'}}(x) = \delta_{x'}(\Phi(x)) = \delta_{x'}(k(x,\cdot)) = k(x,x') = \delta_x(k(\cdot,x')) = [k(\cdot,x')](x),$$

对每个 $x \in X$. 因此，$f_{\delta_{x'}} = k(\cdot,x') \in B_\Phi$. 另一方面，令

$$V_0 := \{\delta_{x'} : x' \in X'\} \subseteq W_0^* = (B_*)^*.$$

于是，

$${}^\perp V_0 := \{g \in B_* : \delta_{x'}(g) = g(x') = 0, \text{ for every } x' \in X'\} = \{g \in B_* : g = 0\} = \{0\}.$$

根据双极定理（Conway，1994，定理 1.8），我们得到

$$B_\Phi = \{0\}^\perp = ({}^\perp V_0)^\perp = \overline{\mathrm{Lin} V_0}^{w*}.$$

因此（3.1）成立，证毕。 □

定义 3.21. 设 $B_\Phi \subseteq \mathcal{F}(X,\mathbb{K})$ 是与特征映射 $\Phi\colon X \to B_*$ 相关联的再生核巴拿赫空间，其中 e

$$B_* := \overline{\mathrm{Lin}\{\Phi(x) : x \in X\}}^{\|\cdot\|}.$$

若其预对偶 $B_* \subseteq \mathcal{F}(X',\mathbb{K})$ 是再生核巴拿赫空间，则称空间 B_Φ 为**特殊特征再生核巴拿赫空间**。

如定理3.20所示，若 B_Φ 是特殊特征再生核巴拿赫空间，则

$$B_\Phi = \overline{\mathrm{Lin}\{k(\cdot,x') : x' \in X'\}}^{w*}.$$

注记 3.22。一个巴拿赫空间 B（具有预对偶 B_*）可能有一个子空间 M，它在 B 中稠密，但不按范数稠密。例如，考虑 $l_\infty \equiv (l_1)^*$，我们有 $c_0{}^{w^*} = l_\infty$，同时 $\overline{c_0}^{\|\cdot\|} = c_0$.。类似地，取 $L_\infty([0,1]) \equiv (L_1([0,1]))^*$，我们有 $C([0,1]))^{\|\cdot\|} = C([0,1]))$，然而 $\overline{C([0,1])}^{w^*} = L_\infty([0,1])$。注意，根据戈尔斯廷定理（Kadets & Iacob，2018年，第 17.2.2 小节定理 1），每个巴拿赫空间在其二次对偶中都是w^*–稠密的。

前一个定理断言，若 B_Φ 是一个特征空间，则对于每个 $x' \in X'$，作为对偶巴拿赫空间的 B_Φ 是包含 $k(\cdot, x')$ 的最小 w^*-闭巴拿赫空间，并且其中对所有 $x \in X$，点求值映射 δ_x 都是连续的。

4 特征再生核巴拿赫空间中的范数插值与正则化

本节将在特征再生核巴拿赫空间的框架下，阐述监督学习问题。从机器学习的角度来看，基于有限数据集进行学习通常被表述为正则化经验风险最小化问题；在插值情形下，则是寻找与观测数据一致的最小范数函数。在再生核希尔伯特空间中，这类表述已得到充分理解，并通过经典表示定理得到有限维解；然而，在巴拿赫空间中，缺乏希尔伯特几何从根本上改变了这一图景。

我们证明，特征再生核巴拿赫空间为函数空间框架提供了一种自然选择，在其中，学习仍可解释为最小范数插值或正则化，同时也明确指出了巴拿赫空间情形下所需的额外结构性假设。我们运用泛函分析论证建立存在性结果，并阐明底层巴拿赫范数的几何如何影响解集。与希尔伯特空间情形不同，表示定理型解不再自动存在。尽管如此，我们给出了这类解存在的充要条件，并由此在再生核巴拿赫空间的背景下得到一般性的表示定理。随后各节将运用这些定理来处理训练过程。

我们强调，特征再生核巴拿赫空间为学习问题表述提供了结构性框架，而有限表示解的存在则需要额外假设，这些假设将在后文给出。

设 $W \subseteq \mathcal{F}(X',\mathbb{K})$ 是一个巴拿赫空间，且 $\Phi\colon X \to W$ 是一个特征映射。假设

$$W = \overline{\mathrm{Lin}\{\Phi(x) : x \in X\}}^{\|\cdot\|},$$

这并不构成限制。设 $B_\Phi \subseteq \mathcal{F}(X,\mathbb{K})$ 是与 Φ 相关联的再生核巴拿赫空间（RKBS），即

$$B_\Phi := \{f_\mu(\cdot) = \mu(\Phi(\cdot)) : \mu \in W^*\},$$

其中 $\|f_\mu\| := \|\mu\|$，对每个 $f_\mu \in B_\Phi$. 与特征映射 Φ 相关联的核映射是函数k： $X \times X' \to \mathbb{K}$，由 $k(x, x') := \delta_{x'}(\Phi(x))$，对每个 $x \in X$ 和 $x' \in X'$给出。这样，空间 B_Φ 也可以用 k表示如下

$$B_\Phi := \{f_\mu(\cdot) = \mu(k(x,\cdot)) : \mu \in W^*\}.$$

给定数据集 $\mathcal{D}_m = \{(x_i, y_i)\}_{i=1}^m \subseteq (X, \mathbb{K})$，令 $\mathcal{L}$： $B_\Phi \to \mathbb{K}^m$ 为对每个 $f_\mu \in B_\Phi$，给出的映射

$$\mathcal{L}(f_\mu) = \big(f_\mu(x_1), \ldots, f_\mu(x_m)\big).$$

令 $\boldsymbol{y} = (y_1, \ldots, y_m) \in \mathbb{K}^m$ 并定义

$$\mathcal{M}_{\boldsymbol{y}} := \{f_\mu \in B_\Phi : \mathcal{L}(f_\mu) = \boldsymbol{y} = (y_1, \ldots, y_m) \in \mathbb{K}^m\}.$$

那么，与数据集 $\mathcal{D}_m$ 相关联的 B_Φ 中的最小范数插值问题如下：

$$\arg\min\{\|f_\mu\| : f_\mu \in \mathcal{M}_{\boldsymbol{y}}\}.$$

在此框架中，这就是所谓的：利用数据集 $\mathcal{D}_m$ 在 B_Φ 中学习一个函数。为求解该问题，定义巴拿赫空间

$$W^{\text{data}} := \text{Lin}\{k(x_1,\cdot), \ldots, k(x_m,\cdot)\} \subseteq W \subseteq \mathcal{F}(X', \mathbb{K}).$$

不失一般性，我们可以假设 $\{k(x_1, \cdot)\ldots\ldots, k(x_m, \cdot)\}$ 线性无关（否则，我们可以使用一个极大的线性无关子集）。数据映射

$$\mu^{\text{data}} : W^{\text{data}} \to \mathbb{K}$$

定义为满足下列条件的唯一线性映射

$$\mu^{\text{data}}(k(x_k,\cdot)) = y_k, \quad \text{for } k = 1, \ldots, m.$$

注意， $f_\mu \in \mathcal{M}_{\boldsymbol{y}}$ 当且仅当 μ 是 μ^{data} 到 W 的延拓（因为 $\mu|_{W^{\text{data}}} = \mu^{\text{data}}$ 意味着，对于每个 $k = 1, \ldots\ldots, m$，都有 $f_\mu(x_k) = y_k$，）。

Hahn-Banach定理（Conway，1994，推论 6.5）保证存在一个从μ^{data} 到 W的哈恩–巴拿赫延拓。这是一个映射

$$\hat{\mu}^{\text{data}} : W \to \mathbb{K}$$

使得

$$\hat{\mu}^{\text{data}}|_{W^{\text{data}}} = \mu^{\text{data}} \quad \text{and} \quad \|\hat{\mu}^{\text{data}}\| = \|\mu^{\text{data}}\|.$$

因此， $f_{\hat{\mu}^{\text{data}}} \in \mathcal{M}_{\boldsymbol{y}}$ with $\|f_{\hat{\mu}^{\text{data}}}\| = \|\mu^{\text{data}}\|$。由于此延拓具有最小范数，我们得到以下结果 .

定理 4.1 (B_Φ中的 MNI 问题)。 设 $W \subseteq \mathcal{F}(X', \mathbb{K})$ 为一个巴拿赫空间，且 Φ： $X \to W$ 是一个映射，使得 $\|\cdot\|W = \text{Lin}\{\Phi(x) : x \in X\}$. 设 k： $X \times X' \to \mathbb{K}$ 为与 Φ 相关联的核，且 $B_\Phi \subseteq \mathcal{F}(X, \mathbb{K})$ 为与 Φ相关联的再生核巴拿赫空间。给定一个数据集

$$\mathcal{D}_m = \{(x_i, y_i)\}_{i=1}^m$$

考虑相应的最小范数插值问题

$$\arg\min\{\|f_\mu\| : f_\mu \in \mathcal{M}_{\boldsymbol{y}}\}. \tag{4.1}$$

那么， $f_\mu \in B_\Phi$ 是(4.1) 的解，当且仅当 μ 是相关数据映射的哈恩–巴拿赫延拓μ^{data}： $W^{\text{data}} := \text{Lin}\{k(x_1,\cdot)\ldots, k(x_m,\cdot)\} \to \mathbb{K}$，定义为

$$\mu^{\text{data}}(k(x_k,\cdot)) = y_k, \quad \textit{for } k = 1, \ldots, m.$$

此外，集合 $\mathcal{S}_{\boldsymbol{y}}$ 的解集(4.1) 非空且为凸集。

根据Hadamard经典判据，一个问题若满足三个条件即为适定：解的存在性、解的唯一性以及稳定性（即解连续依赖于数据，使得微小扰动只引起解的微小变化），参见（Hadamard，1923；Engel和Nagel，2000）。在学习问题中，稳定性往往不成立，因为训练数据中即使很小的扰动也可能导致解发生剧烈变化。因此，这种不适定性要求引入正则化。

为了对其进行正则化 e最小范数插值问题（4.1），通常会在数据保真项上添加一个正则化项，如下 s:

$$\arg\min\{Q_{\boldsymbol{y}}(\mathcal{L}(f_\mu)) + \lambda_0\,\varphi(\|f_\mu\|) : f_\mu \in B_\Phi\} \quad \text{(Regularization Problem)}, \tag{4.2}$$

其中， $Q_{\boldsymbol{y}}$: $\mathbb{K}^m \to \mathbb{R}_0^+$是损失函数（满足 $Q_{\boldsymbol{y}}(\mathbf{y}) = 0$）， φ : $\mathbb{R}^+ \to \mathbb{R}_0^+$是正则化函数， $\lambda_0 > 0$是正则化参数。一般而言， $Q_{\mathbf{y}}$ 和 φ 都是凸且强制的。此外， $Q_{\mathbf{y}}$ 是下半连续的，而 φ 是连续且非递减的。

请注意，正则化问题比 MNI 问题更加一般且灵活，因为它不再要求严格约束 $\mathcal{L}(f_\mu) = \boldsymbol{y}$。这使其能够更好地适用于数据含有噪声或无法精确插值的情形。回想一下，当模型过度学习训练数据（包括不属于潜在模式的噪声或不规则性）时，就会发生**过拟合**。这类模型在未见数据上的表现通常较差，因为它们捕捉到的是伪影，而非一般结构。

正则化问题的解也存在。事实上，若 $f_\mu^0 \in B_\Phi$ 是相关 MNI 问题的一个解，$\arg\min\{\|f_\mu\| : f_\mu \in \mathcal{M}_{\boldsymbol{y}}\}$，则 $\|f_\mu^0\| = \|\mu^{\text{data}}\|$，由此可知，(4.2) 的解的范数不可能大于 $\|\mu^{\text{data}}\|$。因此，若 $B_{(W^{\text{data}})^*}(0,\ \|\mu^{\text{data}}\|)$ 表示以零为中心、半径为 $\|\mu^{\text{data}}\|$ 的 $(W_{\text{数据}})^*$ 闭球，那么函数

$$\mathcal{R}_{\boldsymbol{y}} : \overline{B_{(W^{\text{data}})^*}}\left(0, \|\mu^{\text{data}}\|\right) \to \mathbb{R}_0^+$$

给出

$$\mathcal{R}_{\boldsymbol{y}}(\widetilde{\mu}) = Q_{\boldsymbol{y}}((\widetilde{\mu}(k(x_1,\cdot)), \cdots, \widetilde{\mu}(k(x_m,\cdot))) + \lambda_0\,\varphi(\|\widetilde{\mu}\|)\,,$$

对于每个$\widetilde{\mu} \in \overline{B_{(W^{\text{data}})^*}(0,\ \|\mu^{\text{data}}\|)}$ 都能取得其最小值（Pardalos、Migdalas 和 Pitsoulis，2008，引理 2，第 62 页）。若$\widetilde{\mu}_0$: $W^{\text{data}} \to \mathbb{K}$ 是一个使 $\mathcal{R}_{\mathbf{y}}$ 取得其最小值的泛函，那么 μ_0 的任意哈恩–巴拿赫延拓都是问题（4.2）的一个解。因此，有如下结果：

定理 4.2。设 $W \subseteq \mathcal{F}(X',\ \mathbb{K})$ 是一个巴拿赫空间，且 Φ : $X \to W$ 其中 $W = \overline{\text{Lin}\{\Phi(x) : x \in X\}}$. 令k : $X \times X' \to \mathbb{K}$ 为与 Φ 相关联的核，且 $B_\Phi \subseteq \mathcal{F}(X,\ \mathbb{K})$ 为与 Φ 相关联的再生核巴拿赫空间。给定一个数据集$\mathcal{D}_m = \{(x_i,\ y_i)\}_{i=1}^m$，考虑由下式给出的正则化问题

$$\arg\min\{Q_{\boldsymbol{y}}(\mathcal{L}(f_\mu)) + \lambda_0\,\varphi(\|f_\mu\|) : f_\mu \in B_\Phi\}\,. \tag{4.3}$$

定义数据泛函 μ^{data} : $W^{\text{data}} := \text{Lin}\{k(x_1,\cdot)\ldots, k(x_m,\cdot)\} \to \mathbb{K}$，如下：

$$\mu^{\text{data}}(k(x_k,\cdot)) = y_k, \quad \textit{for } k = 1,\ldots,m.$$

那么，由下式给出的函数 $\overline{B_{(W^{\text{data}})^*}}(0,\ \|\mu^{\text{data}}\|) \to \mathbb{R}_0^+$

$$\mathcal{R}_{\boldsymbol{y}}(\widetilde{\mu}) = Q_{\boldsymbol{y}}((\widetilde{\mu}(k(x_1,\cdot)), \cdots, \widetilde{\mu}(k(x_m,\cdot))) + \lambda_0\,\varphi(\|\widetilde{\mu}\|)\,,$$

达到其最小值。此外，最小化问题（*4.3*）的解集由所有满足以下条件的函数组成：$f_{\widetilde{\mu}} \in B_\Phi$ 其中 $\widetilde{\mu} \in W^*$ 是极小化元 $\widetilde{\mu}_0 \in B_{(W^{\text{data}})^*}(0,\ \|\mu^{\text{data}}\|)$ 的 *Hahn-Banach* 延拓，且该极小化元属于 $\mathcal{R}_{\boldsymbol{y}}$。

Hahn–Banach定理保证了所讨论的最小范数插值问题存在解。然而，一般而言，所需延拓的显式构造尚不清楚，只有在某些特定情形下例外（例如在希尔伯特空间中，可借助里斯–弗雷歇定理构造），因为 Hahn-Banach定理仅能保证所需泛函的存在。下一节中，我们将探讨何时可以由 W^{data}. 中的元素构造出一个解

5 特殊特征再生核巴拿赫空间中的表示定理

表示定理在基于核的学习中发挥着核心作用，它将无限维优化问题化为有限维问题。在再生核希尔伯特空间中，这类结果可直接由希尔伯特结构推出，并适用于广泛的正则化学习问题。然而，在巴拿赫空间中，表示定理不再必然成立，其有效性取决于底层函数空间的其他几何性质和对偶性质。

本节将探讨特殊特征再生核巴拿赫空间中的表示定理。我们将表明，当特征再生核巴拿赫空间具有适当的预对偶结构并满足其他假设（例如，确保求值泛函与巴拿赫几何相容的依赖数据集的条件）时，最小范数插值与正则化问题的解具有有限核表示。这些结果准确界定了经典的基于核的学习原理何时能够推广到巴拿赫空间，以及何时在根本上不再成立。

该簇 RKHS 的经典表示定理（Kimeldorf & Wahba，1971年）指出，给定一个数据集 t

$$\mathcal{D}_m := \{(x_i, y_i)\}_{i=1}^{i=m},$$

在再生核希尔伯特空间中，相应 MNI 范数问题的解总是存在且唯一，并且可以写成在训练点处取值的核函数所构成的有限线性组合；也就是说 $f^{\mathrm{sol}}(\cdot) := \sum_{i=1}^{m} \alpha_i k(\cdot,\ x_i)$。

若 $B_\Phi \subseteq \mathcal{F}(X, \mathbb{K})$ 是特殊特征再生核巴拿赫空间，也就是说，它是一个预对偶 $B_* \subseteq \mathcal{F}(X', \mathbb{K})$ 为再生核巴拿赫空间的特征再生核巴拿赫空间，那么根据定理 3.20，我们知道 $k(\cdot, x') \in B_\Phi$ 并且 $\delta_{x'} \in (B_*)^*$，其中 $B_\Phi \equiv (B_*)^*$ 且 $k(\cdot, x') \equiv \delta_{x'}$.事实上，

$$B_\Phi = \overline{\mathrm{Lin}\{k(\cdot, x') : x' \in X'\}}^{w*} \equiv \overline{\mathrm{Lin}\{\delta_{x'} : x' \in X'\}}^{w*} = (B_*)^*.$$

因此，自然会问，在什么条件下， $f_\mu \in B_\Phi$ MNI 问题（4.1）的解具有以下形式

$$f^{\mathrm{sol}}(\cdot) = \sum_{i=1}^{k} \beta_i k(\cdot, x_i').$$

也就是说，要确定方程 (4.1) 的解（已知其存在于空间$B_\Phi = \mathrm{Lin}\{k(\cdot, x') :\ x' \in X'\}^{w*}$中）何时实际上属于 $\mathrm{Lin}\{k(\cdot, x') :\ x' \in X'\}$. 在下面的结果中，我们将刻画这种情况成立的条件。

为简便起见，在本节中我们始终假设基域为 $\mathbb{K} = \mathbb{R}$。（将以下结果推广到复数情形是显而易见的。）

定理 5.1（特征再生核巴拿赫空间中的表示定理）。 设 $B_\Phi \subseteq \mathcal{F}(X, \mathbb{R})$ 是一个特殊特征再生核巴拿赫空间，$k : X \times X' \to \mathbb{R}$ 是相应的核函数。给定数据集 $\mathcal{D}_m := \{(x_i, y_i)\}_{i=1}^{m}$，考虑与之关联的最小范数插值问题

$$\arg\min\{\|f_\mu\| : f_\mu \in \mathcal{M}_y\}. \tag{5.1}$$

该问题存在如下形式的解

$$f^{\mathrm{sol}}(\cdot) := \sum_{i=1}^{k} \beta_i\, k(\cdot, x_i'),$$

当且仅当存在 $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ 满足以下条件时，某些 $x_1', \ldots, x_k' \in X'$ 才成立

$$\left\| \sum_{i=1}^{m} \alpha_i\, k(x_i, \cdot) \right\| \leq 1 \tag{5.2}$$

以及满足以下条件的 $\beta_1, \ldots, \beta_k \in \mathbb{R}$

$$(\alpha_1 \cdots \alpha_m) \begin{pmatrix} k(x_1, x_1') & \cdots & k(x_1, x_k') \\ \vdots & \ddots & \vdots \\ k(x_m, x_1') & \cdots & k(x_m, x_k') \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \|f^{\mathrm{sol}}\|, \tag{5.3}$$

and

$$\begin{pmatrix} k(x_1, x_1') & \cdots & k(x_1, x_k') \\ \vdots & \ddots & \vdots \\ k(x_m, x_1') & \cdots & k(x_m, x_k') \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}. \tag{5.4}$$

此外， $f^{\mathrm{sol}}(\cdot) := \sum_{i=1}^{k} \beta_i\, k(\cdot, x_i')$ 在 $\sum_{i=1}^{m} \alpha_i k(x_i, \cdot)$处取得其范数。

证明。 根据定理 3.20，我们有 $k(\cdot, x_i') = f_{\delta_{x_i'}}$，因为 是特殊特征再生核巴拿赫空间。因此 e,

$$f^{\mathrm{sol}}(\cdot) := \sum_{i=1}^{k} \beta_i k(\cdot, x_i') = \sum_{i=1}^{k} \beta_i f_{\delta_{x_i'}} = f_{\sum_{i=1}^{k} \beta_i \delta_{x_i'}}.$$

如果 $f\mathrm{sol}(\cdot) := \sum_{i=1}^{k} \beta_i k(\cdot, x_i')$ 是（5.1）的一个解，那么对于每个$j = 1 \cdots, m.$， $\sum_{i=1}^{k} \beta_i k(x_k, x_i') = y_k$，因为 $f^{\mathrm{sol}}(x_j) = y_j$。因此，（5.4）成立。此外，根据定理 4.1，泛函 $\sum_{i=1}^{k} \beta_i \delta_{x_i'} \in W^*$ 是数据映射的一个哈恩–巴拿赫延拓

$$\mu^{\mathrm{data}} : W^{\mathrm{data}} := \mathrm{Lin}\{k(x_1, \cdot), \ldots, k(x_m, \cdot)\} \to \mathbb{R},$$

定义 din (4.2)。因此，存在 $v_0 := \sum_{i=1}^m \alpha_i k(x_i, \cdot) \in W^{\text{data}}$，使得 $\|v_0\| \le 1$ 且

$$\sum_{i=1}^k \beta_i \delta_{x_i'}(v_0) = \left|\sum_{i=1}^k \beta_i \delta_{x_i'}(v_0)\right| = \left\|\sum_{i=1}^k \beta_i \delta_{x_i'}\right\| = \left\|f^{\text{sol}}\right\|.$$

由于 $\sum_{i=1}^k \beta_i \delta_{x_i'}(v_0) = \sum_{i=1}^k \beta_i \delta_{x_i'}(\sum_{i=1}^m \alpha_i k(x_i, \cdot)) = \sum_{i,j} \alpha_i \beta_j k(x_i, x_j')$，我们得出结论，（5.3）也成立。

反过来，由（5.4）可知，$\sum_{i=1}^k \beta_i \delta_{x_i'}$ 是 μ^{data}的延拓。同时，由（5.2）和（5.3）可推出，$v_0 := \sum_{i=1}^m \alpha_i k(x_i, \cdot) \in W$数据满足 $\|v_0\| \le 1$和 $\sum_{i=1}^k \beta_i \delta_{x_i'}(v_0) = \left\|\sum_{i=1}^k \beta_i \delta_{x_i'}\right\| = \|f^{\text{sol}}\|$，因此 $\sum_{i=1}^k \beta_i \delta_{x_i'}$ 是 μ^{data}的哈恩–巴拿赫延拓。由定理 4.1，我们得出结论，$f^{\text{sol}}(\cdot) = f_{\sum_{i=1}^k \beta_i \delta_{x_i'}}$ 是最小范数插值问题（5.1）的解，证毕。其余部分显而易见。 □

阐释与启示。 定理 5.1 阐明了即使缺乏希尔伯特空间结构，巴拿赫空间学习问题何时仍具有有限核表示。适当预对偶的存在以及与数据集的相容性，确保最小范数插值的解可以表示为核函数取值的有限线性组合。这一结果说明了特殊特征再生核巴拿赫空间为何支持核方法学习算法，并为后续章节中发展的向量值构造和神经网络构造奠定了结构基础。
这些条件并非证明技巧的产物，而是反映了一般巴拿赫空间中真实存在的几何障碍。

为了找到具有如下形式的解的充分条件

$$f^{\text{sol}}(\cdot) := \sum_{i=1}^k \beta_i k(\cdot, x_i'),$$

对于某些 $x_1', \cdots, x_k' \in X'$，我们引入下一个定义。

定义 5.2。 设 $B_\Phi \subseteq \mathcal{F}(X, \mathbb{R})$ 是一个特征再生核巴拿赫空间，且 $k: X \times X' \to \mathbb{R}$ 是相应的核函数。若存在 $x_1', \cdots, x_m' \in X'$，使得矩阵满足以下条件，则称数据集 $\mathcal{D}_m := \{(x_i, y_i)\}_{i=1}^m$ 是**良好的**

$$(k(x_i, x_j'))_{ij} \in M_{m\times m}(\mathbb{R})$$

是可逆的。

注意，良好数据集是指满足以下条件的 $\mathcal{D}_m := \{(x_i, y_i)\}_{i=1}^m$：存在元素 $x_1', \cdots, x_m' \in X'$，使得 $(W^{\text{data}})^* = (\text{Lin}\{k(x_1, \cdot) \cdots, k(x_m, \cdot)\})^*$ 中的泛函可以由其在 $x_1', \cdots, x_m'$处的求值唯一确定。更确切地说，采用良好数据集的意义如下。

命题 5.3. 设 $B_\Phi \subseteq \mathcal{F}(X, \mathbb{R})$ 是一个特殊特征再生核巴拿赫空间，其相关核函数为 $k: X \times X' \to \mathbb{R}$.给定一个数据集 $\mathcal{D}_m := \{(x_i, y_i)\}_{i=1}^m$，以及任意的 $x_1', \cdots, x_m' \in X'$，则矩阵

$$(k(x_i, x_j'))_{ij} \in M_{m\times m}(\mathbb{R})$$

当且仅当 $\{\delta_{x_1'|W^{\text{数据}}} \cdots, \delta_{x_m'|W^{data}}\}$是.....的一组基时，才是可逆的

$$(W^{\text{data}})^* = (\text{Lin}\{k(x_1, \cdot), \cdots, k(x_m, \cdot)\})^* = (\text{Lin}\{\Phi(x_1), \cdots, \Phi(x_m)\})^*.$$

证明。 如果 $(k(x_i, x_j'))_{ij} \in M_{m\times m}(\mathbb{R})$ 可逆，那么这些向量

$$\{(k(x_1, x_1'), \cdots, k(x_1, x_m')), \cdots, (k(x_m, x_1'), \cdots, k(x_m, x_m'))\}$$

是线性无关的。因此，函数 $\{k(x_1, \cdot) \cdots, k(x_m, \cdot)\}$ 也是线性无关的，从而它们构成 W^{data}. 的一个基。此外，如果 $\alpha_i \in \mathbb{R}$，对于 $i = 1、\cdots、m$，则有 $\psi := \alpha_1 \delta_{x_1'|W^{\text{data}}} + \cdots + \alpha_m \delta_{x_m'|W^{\text{data}}} = 0$ 当且仅当 $\psi(k(x_k, \cdot)) = 0$，对于 $k = 1、\cdots、m.$。这意味着 $(k(x_i, x_j'))_{ij}\boldsymbol{\alpha}^T = 0$，其中 $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_m) \in M_{1\times m}(\mathbb{R})$。由于 $(k(x_i, x_j'))^{ij}$ 可逆，因此可推出 $\boldsymbol{\alpha} = 0$，从而 $\{\delta_{x_1'|W}\text{data} \cdots、 \delta_{x_m'|W^{\text{data}}}\}$ 线性无关，因此构成 $(W\text{data})^*$. 的一个基

反之，如果 $\{\delta_{x_1'|W^{\text{data}}}, \cdots, \delta_{x_m'|W^{\text{data}}}\}$ 是 $(W^{\text{数据}})^*$ 的基，那么 $\dim((W^{data})^*) = m$，因此 $\dim(W^{\text{data}}) = m$。此外，

$$W^{\text{data}} = \text{Lin}\{k(x_1, \cdot), \cdots, k(x_m, \cdot)\}$$

使得 $\{k(x_1, \cdot) \cdots, k(x_m, \cdot)\}$ 构成一个基。若 $(k(x_i, x_j'))_{ij}$ 不可逆，则存在一个非零的 $\boldsymbol{\alpha} \in M_{1\times m}(\mathbb{R})$，使得 $(k(x_i, x_j'))\boldsymbol{\alpha}^T = 0$。这意味着，对于每个 $j = 1, \cdots, m,$ 都有 $\sum_{i=1}^m \alpha_i \delta_{x_i'|W^{\text{data}}}(k(x_j, \cdot)) = 0,$，因此 $\sum_{i=1}^m \alpha_i \delta_{x_i'|W^{\text{data}}} = 0,$，矛盾。 □

如果数据集 $\mathcal{D}_m := \{(x_i, y_i)\}_{i=1}^{i=m}$是良好的，那么由 $\mu^{\text{data}}(k(x_k,\cdot)) = y_k$, 定义的函数 $\mu^{\text{data}} \in (W^{\text{data}})^*$，对于$k = 1,\cdots, m,$，满足：对于唯一的 $\boldsymbol{\beta} := (\beta_1 \cdots \beta_m) \in M_{1\times m}(\mathbb{R})$，有 $\mu^{\text{data}} := \sum_{i=1}^m \beta_i \delta_{x_i'|W^{\text{data}}}$。因此，探讨由以下表达式给出的 μ^{data} 的延拓何时成立是有意义的

$$\widehat{\mu}^{\text{data}} := \sum_{i=1}^m \beta_i \delta_{x_i'} : W := \overline{\text{Lin}\{k(x,\cdot) : x \in X\}}^{\|\cdot\|} \to \mathbb{R}$$

是 μ^{data} 的哈恩–巴拿赫延拓，因此，我们所讨论的最小范数插值问题的解是否为 $f_{\widehat{\mu}^{\text{data}}}$。显然，只要 $\left\|\mu^{\text{data}}\right\|$ 是延拓 $\widehat{\mu}^{\text{data}}$ 可能具有的最大范数，情况便是如此，下面将对此加以说明。

推论 *5.4.* 设 $B_\Phi \subseteq \mathcal{F}(X,\mathbb{R})$ 是一个特殊特征再生核巴拿赫空间，$k: X \times X' \to \mathbb{R}$ 是相应的核函数。设 $\mathcal{D}_m := \{(x_i, y_i)\}_{i=1}^m$ 是一个良好数据集，并设 $x_1', \cdots, x_m' \in X'$ 满足矩阵 $M := (k(x_i, x_j'))_{ij} \in M_{m\times m}(\mathbb{R})$ 可逆。若 $\boldsymbol{\beta} := (\beta_1 \ldots \beta_m) \in M_{1\times m}(\mathbb{R})$ 满足 $\boldsymbol{\beta}^T := M^{-1}\boldsymbol{y}^T$，其中 $\boldsymbol{y} = (y_1, \cdots, y_m)$，则该问题的解集

$$\arg\min\{\|f_\mu\| : f_\mu \in \mathcal{M}_y\} \tag{5.5}$$

是 $\mathcal{S}_y := \{f_\mu \in B_\Phi : \ \mu$ 是 $\mu^{\text{data}} := \sum_{i=1}^m \beta_i \delta_{x_i'|W^{data}}\}$.的哈恩–巴拿赫延拓

因此，若 $\left\|\mu^{\text{data}}\right\| = \sum_{i=1}^m |\beta_i| \left\|\delta_{x_i'}\right\|$，则 $f^{\text{sol}}(\cdot) := \sum_{i=1}^m \beta_i k(\cdot, x_i')$ 是方程 *(5.5)* 的一个解。

证明。令 $\mu^{\text{data}} \in (W^{\text{data}})^*$ 为由 $\mu^{\text{data}}(k(x_k,\cdot)) = y_k$ 给出的映射，其中 $k = 1, \cdots, m.$ 由命题 5.3 可知，对于唯一的向量 $\boldsymbol{\beta} := (\beta_1 \ldots \beta_m)$，$\mu^{\text{data}} := \sum_{i=1}^m \beta_i \delta_{x_i'|W^{\text{data}}}$ 成立，且该向量属于 $M_{1\times m}(\mathbb{R})$。显然，$\boldsymbol{\beta}$ 由等式 $\boldsymbol{\beta}^T := M^{-1}\boldsymbol{y}^T$. 确定
根据定理 $k(\cdot, x_i') = f_{\delta_{x_i'}}$ 3.20，若 $f^{\text{sol}}(\cdot) := \sum_{i=1}^m \beta_i k(\cdot, x_i')$，则

$$\left\|f^{\text{sol}}\right\| = \left\|f_{\sum_{i=1}^m \beta_i \delta_{x_i'}}\right\| = \left\|\sum_{i=1}^m \beta_i \delta_{x_i'}\right\| \le \sum_{i=1}^m |\beta_i| \left\|\delta_{x_i'}\right\| = \left\|\mu^{\text{data}}\right\|,$$

使得 $\left\|f_{\text{sol}}\right\| = \left\|\mu^{\text{data}}\right\|$。这证明了 f^{sol} 是方程（5.5）的一个解（根据定理 4.1）。 □

上述结果有助于描述我们将在本文最后一节使用的算法。事实上，该算法将基于下一个推论。

对任意 $\alpha \in \mathbb{R}$，定义符号$(\alpha) := +1$若 $\alpha > 0$，符号$(\alpha) := -1$ 若 $\alpha < 0$，且符号$(0) := 0$。
若 $\boldsymbol{\alpha} := (\alpha_1, \cdots, \alpha_m) \in M_{1\times m}(\mathbb{R})$，定义

$$\text{sign}(\boldsymbol{\alpha}) := (\text{sign}(\alpha_1), \cdots, \text{sign}(\alpha_m)) \in \Delta^m,$$

其中 $\Delta := \{-1, 0, 1\}$ 和 $\Delta^m := \Delta \times \overset{m}{\cdots} \times \Delta$.

定义 5.5。设 $B_\Phi \subseteq \mathcal{F}(X, \mathbb{R})$ 是一个特征再生核巴拿赫空间，且 $\mathcal{D}_m := \{(x_i, y_i)\}_{i=1}^m$ 是一个数据集。若符号向量 $\boldsymbol{s} := (s_1, \cdots, s_m) \in \Delta^m$ 满足以下条件，则称其为**可容许**的：

$$\left\|\sum_{i=1}^m s_i k(x_i,\cdot)\right\| \le 1.$$

推论 5.6。设 $B_\Phi \subseteq \mathcal{F}(X,\mathbb{R})$ 是一个特殊特征再生核巴拿赫空间，$k: X \times X' \to \mathbb{R}$ 是对应的核函数。给定一个良好数据集 $\mathcal{D}_m := \{(x_i, y_i)\}_{i=1}^m$，设 $x_1', \cdots, x_m' \in X'$ 满足矩阵$M := (k(x_i, x_j'))_{ij} \in M_{m\times m}(\mathbb{R})$ 为对角矩阵，并且满足

$$|k(x_i, x_i')| = \|k(\cdot, x_i'))\|, \ \textit{for every} \ i \in \{1, \ldots, m\}.$$

设 $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_m) \in M_{1\times m}(\mathbb{R})$ 满足 $\boldsymbol{\beta}^T := M^{-1}\boldsymbol{y}^T$，其中 $\boldsymbol{y} = (y_1, \cdots, y_m)$。若符号$(\boldsymbol{y})$ 是可容许的，则 $f^{\text{sol}}(\cdot) := \sum_{i=1}^m \beta_i k(\cdot, x_i')$ 是以下问题的一个解

$$\arg\min\{\|f_\mu\| : f_\mu \in \mathcal{M}_y\}$$

和 $\left\|f^{\text{sol}}\right\| = \sum_{i=1}^m |\beta_i| \left\|k(\cdot, x_i')\right\|$.

证明。 如果 sign($\boldsymbol{y}$) $=(s_1, \cdots, s_m) \in \Delta^m$, 那么 $\|\sum_{i=1}^m s_i k(x_i, \cdot)\| \leq 1$, 因为 sign($\boldsymbol{y}$) 是可容许的。此外, f^{sol} 在 $\sum_{i=1}^m s_i k(x_i, \cdot)$ 中取得其范数，因为

$$\boldsymbol{\alpha} M \boldsymbol{\beta}^T = \sum_{i=1}^m s_i \left|k(x_i, x_i')\right| \beta_i = \sum_{i=1}^m |\beta_i| \left\|k(\cdot, x_i'))\right\| = \left\|f^{\mathrm{sol}}\right\|.$$

因此，定理5.1 适用，其中 $\boldsymbol{\alpha}=$ 符号($\boldsymbol{y}$) $=(s_1, \cdots, s_m)$，证明完毕。 □

6 向量值再生核巴拿赫空间

许多现代机器学习问题涉及向量值输出，包括多任务学习、结构化预测和多输出神经网络。虽然向量值再生核希尔伯特空间为这类问题提供了成熟的框架，但其巴拿赫空间对应物仍未得到充分探索。本节将特征再生核巴拿赫空间理论扩展到向量值函数空间，从而能够在非希尔伯特范数下研究多输出学习问题。

我们引入向量值特征再生核巴拿赫空间，并发展相关核构造，展示它们如何自然地将学习问题分解为逐坐标分量，同时保留统一的泛函分析结构。这一表述为在基于核的巴拿赫空间框架内理解多输出模型（包括具有多个输出的神经网络）奠定了数学基础。

为此，我们考虑巴拿赫空间 $B \subseteq \mathcal{F}(X, \mathbb{K}^m)$，其中

$$\mathcal{F}(X, \mathbb{K}^m) := \{f : X \to \mathbb{K}^m : f \text{ is a function}\}.$$

通常，这意味着巴拿赫空间 B 是 $\mathcal{F}(X, \mathbb{K}^m)$ 的一个向量子空间（因此 $\mathbb{K}$ 是 B的基域）。

记 $\mathcal{L}(B, \mathbb{K}^m)$ 为从 B 到 $\mathbb{K}^m$. 的所有线性映射构成的映射。根据（Lin、Song 和 Zhang, 2021, 定义 2.1），有如下定义。

定义 6.1。向量值再生核巴拿赫空间 是一个巴拿赫空间 $B \subseteq \mathcal{F}(X, \mathbb{K}^m)$，满足对于每个 $x \in X.$，由 $\delta_x(f) = f(x)$ 给出的映射 $\delta\colon X \to \mathcal{L}(B, \mathbb{K}^m)$ 是连续的

众所周知 ，线性映射 $\delta_x\colon B \to \mathbb{K}^m$ 连续，当且仅当存在 $C_x \geq 0$ 使得 at

$$|\pi_i \delta_x(f)| := |\pi_i f(x)| \leq C_x \|f\|, \text{ for every } f \in B \text{ and every } i = 1, \ldots, m.$$

其中 π_i： $\mathbb{K}^m \to \mathbb{K}$ 是对第 i 个坐标的典范投影。换言之，对于每个 $i = 1$, ..., $m.$, δ_x 的连续性等价于 $\pi_i \delta_x$ 的连续性。因此，我们得到以下结果。

命题 6.2。 设 $B \subseteq \mathcal{F}(X, \mathbb{K}^m)$ 是一个巴拿赫空间，并且

$$B_i := \{\pi_i f : X \to \mathbb{K} : f \in B\} \subseteq \mathcal{F}(X, \mathbb{K}), \quad (i = 1, \ldots, m).$$

当且仅当对于每个 $i = 1$, ..., m, B_i 都是再生核巴拿赫空间时， B 才是向量值再生核巴拿赫空间。

设 $B \subseteq \mathcal{F}(X, \mathbb{K}^m)$ 是一个向量值再生核巴拿赫空间。这意味着 $\pi_i \delta_x \in B^*$, 对于每个 $i = 1$, ..., $m.$ 可能的原因是 $\pi_i \delta_x \in J(w_x^i)$ 对于某个 $w_x^i \in B_*$, 对于每个 $i = 1$, ..., $m.$ 如我们已经证明的（见命题 3.13），w_x^i 必定唯一，因此存在一个函数 Φ_i： $X \to B_*$，使得 $J_{\Phi_i(x)} \equiv \pi_i \delta_x$。由此引出以下定义。

定义 6.3。 设 $B \subseteq \mathcal{F}(X, \mathbb{K}^m)$ 为一个具有预对偶 $B_* \subseteq \mathcal{F}(X', \mathbb{K})$ 的巴拿赫空间，，满足 $(B_*)^* \equiv B$。(B, B_*) 上的**向量值特征映射**是一个映射

$$\Phi := (\Phi_1, \ldots, \Phi_m),$$

其中 Φ_i： $X \to B_*$，对于每个 $i = 1$, ..., m，满足

$$B_* = \overline{\mathrm{Lin}\{\Phi_i(x) : x \in X,\ i = 1, \ldots, m\}}^{\|\cdot\|}.$$

与 $\Phi =$ （Φ_1, $\cdots \Phi_m$）相关联的向量值特征再生核巴拿赫空间是

$$B_\Phi := \{f_\mu : \mu \in (B_*)^*\} \subseteq \mathcal{F}(X, \mathbb{K}^m), \tag{6.1}$$

其中，对于每个 $x \in X$,

$$f_\mu(x) := \mu(\Phi(x)) = (\mu(\Phi_1(x)), \cdots, \mu(\Phi_m(x))) \in \mathbb{K}^m,$$

是一个赋有范数 $\|f_\mu\| := \|\mu\|$. 的巴拿赫空间。事实上， $B_\Phi \equiv (B_*)^* \equiv B$.

注 6.4. 根据定义 3.6，在标量值情形下，特征映射是映射 $\Phi\colon X \to W \subseteq \mathcal{F}(X', \mathbb{K})$，其中 X 和 X' 是非空集合，W 是一个巴拿赫空间。此外，与 Φ 相关联的再生核巴拿赫空间是例 3.5 中描述的空间 B_Φ，而 $B_* := \text{Lin}\{\Phi(x) : x \in X\}^{\|\cdot\|}$ 是 B_Φ. 的预对偶。因此，根据定义 6.1，更准确地说，我们可以说 Φ 是 (B_Φ, B_*) 上的特征映射。

正如下面所示，每个向量值特征映射都关联有一个核函数。

定义 6.5。设 $B \subseteq \mathcal{F}(X, \mathbb{K}^m)$ 是一个具有预对偶 $B_* \subseteq \mathcal{F}(X', \mathbb{K})$ 的巴拿赫空间。一个**向量值核函数定义在** (B, B_*) 上，是一个映射

$$k := (k_1, \ldots, k_m) : X \times X' \to \mathbb{K}^m$$

使得

$$B_* = \overline{\text{Lin}\{k_i(x, \cdot) : x \in X,\ i = 1, \ldots, m\}}^{\|\cdot\|}$$

并且 $k_i(x, x') = \delta_{x'}(k_i(x, \cdot))$，对于任意 $x \in X$，$x' \in X'$，以及 $i = 1, \ldots, m$.

下一个结果的证明是显然的。

命题 6.6. 设 $B \subseteq \mathcal{F}(X, \mathbb{K}^m)$ 是一个巴拿赫空间，其预对偶为 $B_* \subseteq \mathcal{F}(X', \mathbb{K})$。每个定义在 (B, B_*) 上、取值于 $=(\Phi_1, \ldots, \Phi_m)$ 的向量值特征映射 Φ，都唯一对应于定义在 (B, B_*) 上的一个向量值核函数，记作 $k = (k_1, \ldots, k_m) : X \times X' \to \mathbb{K}^m$；反之亦然，并满足：对每个 $i = 1, \ldots, m$，$k_i(x, \cdot) := \Phi_i(x)$。

设 k 是定义在 (B, B_*) 上的向量值核函数。前一个结果使我们能够用相关核来表示空间 B_Φ，该空间定义于 (6.1) 中。也就是说，考虑如下空间

$$B_\Phi^k := \{\widehat{f}_\mu : \mu \in (B_*)^*\} \subseteq \mathcal{F}(X, \mathbb{K}^m), \tag{6.2}$$

其中，对每个 $x \in X$,

$$\widehat{f}_\mu(x) := \mu(k(x, \cdot)) = (\mu(k_1(x)), \cdots, \mu(k_m(x, \cdot))) \in \mathbb{K}^m.$$

我们称 B_Φ^k 为**与向量值核函数** k. 相关的再生核巴拿赫空间。显然，B_Φ^k 是一个配备范数 $\left\|\widehat{f_\mu}\right\| := \|\mu\|$. 的巴拿赫空间。事实上，在如下识别下，$B_\Phi = B_\Phi^k \equiv (B_*)^* \equiv B$

$$f_\mu = \widehat{f}_\mu \equiv \mu \equiv \varphi(\mu),$$

当 $(B_*)^* \overset{\varphi}{\equiv} B$.

以下结果有助于刻画向量值特征再生核Banach空间。此类空间的定义将在后文介绍。

定理 6.7。设 $B \subseteq \mathcal{F}(X, \mathbb{K}^m)$ 是一个巴拿赫空间，且设 $B_* \subseteq \mathcal{F}(X', \mathbb{K}^m)$ 是 B 的预对偶，其中 $B \overset{\varphi}{\equiv} (B_*)^*$。那么，以下断言等价。

(i) 对于每个 $x \in X$ 以及每个 $i = 1, \ldots, m$，存在 $w_x^i \in B_*$ 使得 $\pi_i \delta_x \overset{\varphi^*}{\equiv} J_{w_x^i} \in J(B_*)$。

(ii) 对于每个 $i = 1, \ldots, m$，存在一个映射 $\Phi_i\colon X \to B_*$，使得 $J_{\Phi_i(x)} \overset{\varphi^*}{\equiv} \pi_i \delta_x$，对每个 $x \in X$.

(iii) 存在一个向量值特征映射 $\Phi := (\Phi_1, \ldots, \Phi_m)$ 定义在(B, B_*)上，使得 $B = B_\Phi$.

(iv) 存在 $k = (k_1, \ldots, k_m)\colon X \times X' \to \mathbb{K}^m$，使得 $k_i(x, \cdot) \in B_*$，并且 $\pi_i \delta_x \overset{\varphi^*}{\equiv} J_{k_i(x, \cdot)}$，对每一个 $x \in X$ 和 $i = 1, \ldots, m$。

(v) 存在 一个定义在(B, B_*)上的向量值核函数，记作 $k\colon X \times X' \to \mathbb{K}^m$，使得 $B = B_\Phi^k$.

如果上述断言均成立，那么 Φ和 k 是唯一的。

证明。 (i) $\iff$ (ii) 该断言可直接由命题 3.13得出。

(ii) $\implies$ (iii)。令 $\mu \in (B_*)^*$ 满足：对每个 $i = 1, \ldots, m$，以及每个 $x \in X$，$\mu(\Phi_i(x)) = 0$。因此，$J_{\Phi_i(x)}(\mu) = 0$。令 $f \in B$ 满足 $\varphi(f) = \mu$，并令 $x \in X$。由于 $J_{\Phi_i(x)} \overset{\varphi^*}{\equiv} \pi_i \delta_x$，我们有

$$J_{\Phi_i(x)}(\mu) = J_{\Phi_i(x)}(\varphi(f)) = \pi_i \delta_x(f) = 0, \quad \text{for all } i = 1, \cdots m,$$

so t帽 $f(x) = 0$。由此可得 $f = 0$ 因此 $\mu = 0$。因此，根据（Conway，1994，推论 6.14 ），

$$B_* := \overline{\mathrm{Lin}\{\Phi_i(x) : x \in X,\ i = 1, \ldots, m\}}^{\|\cdot\|}.$$

这证明了 Φ 是定义在 (B, B_*) 上的向量值特征映射。 $B = B_\Phi$ 由以下事实可得：对于每个 $x \in X$ 和每个 $f \in B$，$f(x) = f_{\varphi(f)}(x)$。事实上，

$$\begin{aligned} f_{\varphi(f)}(x) &= ([\varphi(f)](\Phi_1(x)), \ldots, [\varphi(f)](\Phi_m(x))) = (J_{\Phi_1(x)}(\varphi(f)), \ldots, J_{\Phi_m(x)}(\varphi(f))) \\ &= (\pi_1\delta_x(f), \ldots, \pi_m\delta_x(f)) = (f_1(x), \ldots, f_m(x)) = f(x). \end{aligned}$$

（iii）$\Longrightarrow$（i）。由于 $B = B_\Phi$，且 $f = f_{\varphi(f)}$，对于每个 $f \in B$，我们有

$$\pi_i\delta_x(f) = \pi_i\delta_x(f_{\varphi(f)}) = J_{\Phi_i(x)}(\varphi(f)).$$

因此，$\pi_i\delta_x = J_{\Phi_i(x)}\varphi$，从而对于每个 $x \in X$ 和 $i = 1$，……，m，都有$\pi_i\delta_x \equiv J_{\Phi_i(x)}$。

等价关系（ii）$\Longrightarrow$（iv）以及（iii）$\Longrightarrow$（v）由命题 6.6得出。

使用定理 Φ 中给出的相同论证，可以清楚地得出 k 和3.16的唯一性。

□

定义 6.8。设 $B \subseteq \mathcal{F}(X, \mathbb{K}^m)$ 是一个巴拿赫空间，其预对偶为 $B_* \subseteq \mathcal{F}(X', \mathbb{K})$。若存在定义在 (B, B_*) 上的向量值特征映射 $\Phi := (\Phi_1, \ldots, \Phi_m)$，使得 $B = B_\Phi$，则称 B 是一个**向量值特征再生核巴拿赫空间**。

显然，定理6.7 中陈述的每一个等价断言都给出了向量值特征再生核巴拿赫空间这一概念的一种刻画。

定理 6.9。设 $B \subseteq \mathcal{F}(X, \mathbb{K}^m)$ 是一个向量值特征再生核巴拿赫空间，$B_* \subseteq \mathcal{F}(X', \mathbb{K})$ 是 B 的预对偶，且 $\Phi := (\Phi_1, \ldots, \Phi_m)$ 是定义在 (B, B_*) 上的向量值特征映射，使得 $B = B_\Phi$。若 B_* 是一个再生核巴拿赫空间，则 $k(\cdot, x') \in B, \delta'_x \in (B_*)^*$，且 $f_{\delta_{x'}} = k(\cdot, x')$。此外，

$$\overline{\mathrm{Lin}\{k(\cdot, x') : x' \in X'\}}^{w^*} = B.$$

证明。如果 B_* 是一个再生核巴拿赫空间，则对于每个 $x' \in X'$，有 $\delta_{x'} \in (B_*)^*$，因此 $f_{\delta_{x'}} \in B_\Phi = B$。如果 $x \in X$，则

$$f_{\delta_{x'}}(x) := \delta_{x'}(k(x, \cdot)) = k(\cdot, x')(x).$$

从而对于每个 $x' \in X'$，有 $k(\cdot, x') = f_{\delta_{x'}}$。令 $V_0 := \{\delta_{x'} : x' \in X'\} \subseteq (B_*)^*$。于是，

$${}^\perp V_0 = \{g \in B_* : \delta_{x'}(g) = 0,\ x' \in X'\} = \{g \in B_* : g(x') = 0,\ x' \in X'\} = \{0\}.$$

根据双极定理（Conway，1994，1.8. 双极定理），我们得到

$$({}^\perp V_0)^\perp = \overline{\mathrm{Lin}\{V_0\}}^{w^*} = (B_*)^*.$$

由于 $\delta_{x'} \equiv f_{\delta_{x'}} = k(\cdot, x')$，且 $(B_*)^* \equiv B = B_\Phi$，因此

$$\overline{\mathrm{Lin}\{k(\cdot, x') : x' \in X'\}}^{w^*} = B,$$

如所要求。 □

定义 6.10。设 $B = B_\Phi \subseteq \mathcal{F}(X, \mathbb{K}^m)$ 是一个向量值特征再生核巴拿赫空间，与定义在 (B, B_*) 上的向量值特征映射 $\Phi := (\Phi_1, \ldots, \Phi_m)$ 相关联，其中巴拿赫空间 $B_* \subseteq \mathcal{F}(X', \mathbb{K})$ 是 B 的预对偶。我们称 B 为**特殊向量值再生核巴拿赫空间**，如果 B_* 是一个再生核巴拿赫空间。因此，

$$\overline{\mathrm{Lin}\{k(\cdot, x') : x' \in X'\}}^{w^*} = B = B_\Phi.$$

注释 6.11。若 $W_1, \ldots, W_m$ 是巴拿赫空间，则 $W_1 \times \cdots \times W_m$ 配备该范数后成为一个巴拿赫空间

$$\|(w_1, \ldots, w_m)\|_\infty = \max\{\|w_1\|, \ldots, \|w_m\|\}, \quad (w_1, \ldots, w_m) \in W \times \cdots \times W.$$

我们将这个巴拿赫空间记作 $W_1 \times {}^{\|\cdot\|_\infty} \times W_m$. 该空间的拓扑对偶，即 $(W_1 \times {}^{\|\cdot\|_\infty} \times W_m)^*$，与 $W_1^* \times {}^{\|\cdot\|_1} \times W_m^*$，等距同构，也就是赋予范数的空间 $W_1^* \times \cdots \times W_m^*$

$$\|(\mu_1, \ldots, \mu_m)\|_1 = \sum_{i=1}^m \|\mu_i\|, \text{ for every } (\mu_1, \ldots, \mu_m) \in W_1^* \times \cdots \times W_m^*.$$

换言之，$(W_1 \times {}^{\|\cdot\|_\infty}\times, W_m)^* \equiv W_1^* \times {}^{\|\cdot\|_1}\times W_m^*$，其中等距同构 $\varphi: W_1^* \times {}^{\|\cdot\|_1}\times W_m^* \to (W_1 \times {}^{\|\cdot\|_\infty}\times W_m)^*$ 由下式给出

$$[\varphi(\mu_1,\ldots,\mu_m)](w_1,\ldots,w_m) := \sum_{i=1}^m \mu_i(w_i) \in \mathbb{K}.$$

同样，$(W_1 \times {}^{\|\cdot\|_\infty}\times W_m)^{**} \equiv W_1^{**} \times {}^{\|\cdot\|_\infty}\times W_m^{**}$（Megginson，1998，定理 1.10.13）。在这种情况下，等距同构为 ψ：$W_1^{**} \times {}^{\|\cdot\|_\infty}\times W_m^{**} \to (W_1 \times {}^{\|\cdot\|_\infty}\times W_m)^{**}$ 由下式给出

$$[\psi(w_1^{**},\ldots,w_m^{**})](\mu) := \sum_{i=1}^m w_i^{**}(\mu_i) \in \mathbb{K},$$

对于每个 $\mu \in (W_1 \times {}^{\|\cdot\|_\infty}\times W_m)^*$ 以及每个 $(w_1^{**}, \ldots\ldots, w_m^{**}) \in W_1^{**} \times {}^{\|\cdot\|_\infty}\times W_m^{**}$，只要 $\mu \equiv (\mu_1, \ldots\ldots, \mu_m)$。最后注意，典范嵌入

$$J: W_1 \times {}^{\|\cdot\|_\infty}\times W_m \to (W_1 \times {}^{\|\cdot\|_\infty}\times W_m)^{**} \equiv W_1^{**} \times {}^{\|\cdot\|_\infty}\times W_m^{**}$$

恰好是

$$J \equiv (J^1,\ldots,J^m),$$

其中 $J^i: W_i \to W_i^{**}$ 是与 W_i. 相关联的典范嵌入。这意味着，若 $\mu \equiv (\mu_1, \ldots\ldots, \mu_m)$，则，

$$J_{(w_1,\ldots,w_m)}(\mu) := (J^1_{w_1},\ldots,J^m_{w_m})(\mu_1,\ldots,\mu_m) = \sum_{i=1}^m J^i_{w_i}(\mu_i) = \sum_{i=1}^m \mu_i(w_i) \in \mathbb{K},$$

对于每个 $(w_1, \ldots\ldots, w_m) \in W_1 \times {}^{\|\cdot\|_\infty}\times W_m$. 因此，

$$J_{(0,\ldots w_i,\ldots,0)}(\mu) = J^i_{w_i}(\mu_i) = \mu_i(w_i) \in \mathbb{K}.$$

如下面的例子所示，构造向量值特征再生核巴拿赫空间的例子相当直接。

例 6.12。(向量值特征再生核Banach空间原型)。设 X 是一个非空集合，$W_i \subseteq \mathcal{F}(X', \mathbb{K})$ 是一个巴拿赫空间，且 $\Phi_i^0: X \to W_i$ 是一个特征映射，对每个 $i = 1, \ldots\ldots, m$. 不失一般性，可假设

$$W_i := \overline{\mathrm{Lin}\{\Phi_i^0(x) : x \in X\}}^{\|\cdot\|}.$$

现在我们定义

$$B_{\Phi^0} := \{f_\mu(\cdot) : \mu \equiv (\mu_1,\ldots,\mu_m) \in W_1^* \times {}^{\|\cdot\|_1}\times W_m^*\} \subseteq \mathcal{F}(X, \mathbb{K}^m),$$

其中，对每个 $x \in X$，

$$f_\mu(x) := \mu(\Phi^0(x)) = (\mu_1(\Phi_1^0(x)),\ldots,\mu_m(\Phi_m^0(x))),$$

赋予范数

$$\|f_\mu\| := \sum_{i=1}^m \left\|\mu_i + W_i^\perp\right\| = \sum_{i=1}^m \|\mu_{i|W_i}\|.$$

那么，B_{Φ^0} 是一个巴拿赫空间，且 $B_* := W_1 \times {}^{\|\cdot\|_\infty}\times W_m$ 是 B_{Φ^0}. 的预对偶。事实上，考虑赋予以下范数的特征再生核巴拿赫空间

$$B_{\Phi_i^0} := \{\pi_i f_\mu(\cdot) := \mu_i(\Phi_i^0(\cdot)) : \mu_i \in W_i^*\} \subseteq \mathcal{F}(X, \mathbb{K})$$

赋予范数 $\|\pi_i f_\mu\| = \|\mu_i\|$后，我们得到

$$B_{\Phi^0} \equiv B_{\Phi_1^0} \times {}^{\|\cdot\|_1}\times B_{\Phi_m^0}.$$

此外，B_{Φ^0} 是一个向量值特征再生核巴拿赫空间。事实上，映射 $\Phi_i^0: X \to W_i$ 可以视为映射

$$\Phi_i : X \to B_* := W_1 \times {}^{\|\cdot\|_\infty}\times W_m$$

由下式给出

$$\Phi_i(x) := (0,\ldots,\Phi_i^0(x),\ldots 0) \in B_* := W_1 \times {}^{\|\cdot\|_\infty}\times W_m.$$

因此，

$$J\Phi_i(x) = (0, \ldots, J^i\Phi_i^0(x), \ldots 0) \in (B_*)^{**} \equiv W_1^{**} \times {}^{\|\cdot\|_\infty} \times W_m^{**}.$$

根据上述备注，如果 $\mu = (\mu_1, \ldots\ldots, \mu_m) \in W_1^* \times {}^{\|\cdot\|_1} \times W_m^*$，我们有

$$J_{\Phi_i(x)}(\mu) = J_{\Phi_i^0(x)}(\mu_i) = \mu_i(\Phi_i^0(x)) = \pi_i\delta_x(f_\mu) \quad (f_\mu \in B_{\Phi^0}),$$

因此$J_{\Phi_i(x)} \equiv \pi_i\delta_x$，对于每个 $x \in X$，以及每个 $i = 1, \ldots\ldots, m.$，因此，$B_{\Phi^0} = B_\Phi$，这意味着 B_{Φ^0} is 一个向量值特征 再生核巴拿赫空间。注意

$$B_\Phi \equiv (W_1 \times {}^{\|\cdot\|_\infty} \times W_m)^* \equiv W_1^* \times {}^{\|\cdot\|_1} \times W_m^*.$$

因此，$B_\Phi^* \equiv W_1^{**} \times {}^{\|\cdot\|_\infty} \times W_m^{**}$和 $B_* := W_1 \times {}^{\|\cdot\|_\infty} \times W_m$是 B_Φ的预对偶

我们得出结论：给定 m 巴拿赫空间 $W_i \subseteq \mathcal{F}(X, \mathbb{K})$及其相关特征映射 $\Phi_i^0 : X \to W_i$, 满足$W_i := \text{Lin}\{\Phi_i^0(x) : x \in X\}$, 对于每个 $i = 1, \ldots, m,$，空间

$$B_\Phi \equiv B_{\Phi_1^0} \times {}^{\|\cdot\|_1} \times B_{\Phi_m^0} \subseteq \mathcal{F}(X, \mathbb{K}^m)$$

是一个向量值特征再生核巴拿赫空间，我们称之为**向量值特征再生核Banach空间原型**。

此外，如果求值映射 $\delta_{x'} : W_i \to \mathbb{K}$ 对每个 $x' \in X'$和 $i = 1 \ldots, m$ 均连续，那么根据定理6.9，函数 $k(\cdot, x')$ 满足

$$k(\cdot, x') = (\delta_{x'}(\Phi_1^0(\cdot)), ..., \delta_{x'}(\Phi_m^0(\cdot)) \in B_{\Phi^0} = B_\Phi,$$

and

$$\overline{\text{Lin}\{k(\cdot, x') : x' \in X'\}}^{w^*} = B_{\Phi^0} = B_\Phi.$$

7 神经网络作为向量值再生核巴拿赫空间

神经网络是现代机器学习中表达能力最强且应用最广泛的模型之一，但其理论分析往往依赖参数空间论证或无限宽核极限。本节采用一种互补的函数空间视角，说明固定架构神经网络可以自然地嵌入特殊的向量值特征再生核巴拿赫空间。

从这一视角来看，训练神经网络相当于在适当的函数空间中求解最小范数插值问题或正则化问题，其中范数由所选的参数范数诱导。这种解释无需借助无限宽极限，便将核方法与神经网络联系起来，并阐明架构选择和参数范数如何诱导特定的函数空间几何。我们通过具体示例展示这一框架，说明神经网络如何自然地融入前文构建的再生核巴拿赫空间形式体系。尽管我们的分析聚焦于固定架构，该框架也适用于任意有限架构以及参数空间上的范数。

在下文中，我们旨在将前几节所建立的理论框架应用于深度神经网络的情形。出于这一目的，我们引入深度神经网络的概念，沿用（Berlyand & Jabin，2023年，第4章）中的表述，但为了一致性和清晰性，将这些定义调整为我们的记号。

在给出深度神经网络的定义之前，我们需要定义其基本组成部分：神经元函数。

定义 7.1. 一个**神经元函数** $f : \mathbb{R}^s \to \mathbb{R}$ 是如下形式的映射

$$f(\mathbf{x}) = \sigma(\omega \cdot \mathbf{x} + b), \qquad (\mathbf{x} \in \mathbb{R}^s) \tag{7.1}$$

其中 $\sigma : \mathbb{R} \to \mathbb{R}$ 是一个连续的非线性函数，称为**激活函数**，$\omega \in \mathbb{R}^s$ 是一个**权重**向量，而$b \in \mathbb{R}$ 是一个标量，称为**偏置**。这里，$\omega \cdot \mathbf{x}$ 表示 $\mathbb{R}^s$上的内积。

由于人工神经元的输出是一个标量，我们可以将若干神经元组合起来，构造一个向量值函数，称为**层函数**。

定义 7.2. 一个**层函数** $g : \mathbb{R}^s \to \mathbb{R}^n$ 是如下形式的映射

$$g(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_n(\mathbf{x})),$$

其中每个 $f_i : \mathbb{R}^s \to \mathbb{R}$ 是一个**人工神经元**，如式（7.1）所示，并拥有自己的权重向量 $w_i = (w_{i1}, \ldots, w_{is})$ 和偏置b_i，对于每个 $i = 1, \ldots, n$。

注7.3。一般而言，每个神经元 f_i 都可以有自己的激活函数 φ_i。然而，实践中通常假设同一层中的所有神经元共用同一个激活函数，这样可以简化记号和实现。

注7.4。第一层，即输入层，通常不被视为真正的层函数。因此，如果称一个网络具有 l 个层函数，则实际上总共有 $l+1$ 层。在实践中，“层函数”和“层”这两个术语经常互换使用，因此当称一个网络具有 k 层时，并不计入输入层，层的总数为 $k+1$，其中包括隐藏层和输出层。

定义7.2 可以等价地改写为如下矩阵形式

$$g(\mathbf{x}) = \bar{\sigma}(W\mathbf{x} + \boldsymbol{b}), \tag{7.2}$$

其中 $W \in \mathbb{R}^{n\times s}$ 是**权重矩阵**，$\boldsymbol{b} \in \mathbb{R}^n$ 是**偏置向量**，且 $\bar{\sigma}: \mathbb{R}^n \to \mathbb{R}^n$ 是**向量激活函数**定义为

$$\bar{\sigma}(z_1, \ldots, z_n) = \big(\sigma(z_1), \ldots, \sigma(z_n)\big), \tag{7.3}$$

对于标量激活 关于函数 σ，如定义7.1所述。在这一表述中，第 i个预激活值为

$$z_i = w_{i1}x_1 + \cdots + w_{is}x_s + b_i, \qquad i = 1, \ldots, n, \tag{7.4}$$

因此，$g(\mathbf{x})$ 的第 i个分量为 $f_i(\mathbf{x}) = \sigma(z_i)$。

如备注7.4所述，参数 l表示层函数的数量，并决定网络的深度。若 $l = 1$ 或 $l = 2$，则该网络至多有一个隐藏层，称为浅层人工神经网络（ANN）。下文主要关注 $l \geq 3$的情形，即除输入层和输出层外，至少有两个隐藏层的网络。这类架构通常称为**深度神经网络**（DNN）。

深度神经网 络完全由其层宽度以及可训练参数集合确定 s

$$\{(W^{(k)}, b^{(k)}) : k = 1, \ldots, l\},$$

其中，$W^{(k)} \in \mathbb{R}^{m_k \times m_{k-1}}$ 和 $\boldsymbol{b}^{(k)} \in \mathbb{R}^{m_k}$ 分别表示第 k层的权重矩阵和偏置向量。

在此背景下，为了以统一的方式表示所有参数，我们引入参数空间 e

$$\Theta := \prod_{k=1}^{l} [\mathcal{M}_{m_k \times m_{k-1}}(\mathbb{R}) \times \mathcal{M}_{m_k \times 1}(\mathbb{R})],$$

使得每个 $\theta \in \Theta$ 都对应于一个族 $\{(W^{(k)}, \boldsymbol{b}^{(k)})\}_{k=1}^{l}$。

注 7.5。注意到 $m_0 := s$以及 $m_l := t$。

定义 7.6。一个 **神经网络**是一个函数 $f_\theta^{\text{net}}: \mathbb{R}^s \to \mathbb{R}^t$ 其形式为

$$f_\theta^{\text{net}}(\mathbf{x}) = \big(g^{(l)} \circ g^{(l-1)} \circ \cdots \circ g^{(1)}\big)(\mathbf{x}), \quad l \geq 1, \tag{7.5}$$

其中每个 $g^{(i)}: \mathbb{R}^{m_{i-1}} \to \mathbb{R}^{m_i}$ 是第 i 层的层函数（见（7.2）），具有其自身的权重矩阵 $W^{(i)}$和偏置向量 $\boldsymbol{b}^{(i)}$。如果 $l \geq 3$，那么我们称 f_θ^{net}为 **深度神经网络**。

现在，让我们考虑所有具有固定架构的神经网络类，该架构由 l 层、输入层中的 s 个神经元以及输出层中的 t个神经元组成。我们将此类记为

$$\big\{ f_\theta^{\text{net}} : \mathbb{R}^s \to \mathbb{R}^t \;\big|\; \theta \in \Theta \big\} \subseteq \mathcal{F}(\mathbb{R}^s, \mathbb{R}^t),$$

其中 Θ是相应的参数空间。

设 $W_i \subseteq \mathcal{F}(\Theta, \mathbb{R})$ 是一个巴拿赫空间。对于$\mathbf{x} \in \mathbb{R}^s$ 和 $\theta \in \Theta$，我们定义

$$k : \mathbb{R}^s \times \Theta \to \mathbb{R}^t \;\text{ as }\; k(\mathbf{x}, \theta) := f_\theta^{\text{net}}(\mathbf{x})\,\xi(\theta). \tag{7.6}$$

写作 $k = (k_1, \ldots, k_t)$，函数 ξ：$\Theta \to \mathbb{R}^+$需满足：对于每个 $\mathbf{x} \in \mathbb{R}^s$ 以及每个 $i = 1, \ldots, t$，函数 $k_i(\mathbf{x}\ \cdot)$ 都属于 W_i。例如，若取 $W_i = C_0(\Theta)$，即在无穷远处趋于零的连续函数空间，则必须选择 ξ 使得，对于每个 $\mathbf{x} \in \mathbb{R}^s$，

$$\theta \in \Theta \mapsto f_\theta^{\text{net}}(\mathbf{x})\,\xi(\theta) \in \mathbb{R}^t \text{ is continuous and } \lim_{\|\theta\| \to \infty} \big\| f_\theta^{\text{net}}(\mathbf{x})\,\xi(\theta) \big\| = 0,$$

或者等价地，

$$\lim_{\|\theta\|\to\infty} \pi_i f_\theta^{\text{net}}(\mathbf{x})\,\xi(\theta) = 0 \quad \text{for all } i = 1, \ldots, t.$$

令 $W := W_1 \times^{\|\cdot\|_\infty} \cdots \times W_t$ 为赋予乘积范数 $\|(w_1, \ldots, w_t)\|_\infty := \max_{i=1\cdots t} \|w_i$ 的巴拿赫空间 $\|$
并设

$$B_* \;:=\; \overline{\text{Lin}\{\,(k_1(\mathbf{x},\cdot),\ldots,k_t(\mathbf{x},\cdot)) :\; \mathbf{x}\in\mathbb{R}^s\,\}}^{\|\cdot\|_\infty} \;\subseteq\; W, \;\text{ and }\; B \equiv (B_*)^*.$$

根据这些定义， k 成为定义 (B, B_*) 意义下的向量值核函数，参见定义 6.5。等价地，对于每个 $i = 1, \ldots, t$，令

$$W_i := \overline{\text{Lin}\{\,k_i(\mathbf{x},\cdot) :\; \mathbf{x}\in\mathbb{R}^s\,\}}^{\|\cdot\|_\infty} \quad \text{and let} \quad B_{\Phi_i}^k \equiv (W_i)^*.$$

然后，一个向量值 与核函数（7.6）相关联的再生核巴拿赫空间是空间 B_Φ^k 由乘积给出 t

$$B_\Phi^k \;:=\; B_{\Phi_1}^k \times^{\|\cdot\|_1} \cdots \times B_{\Phi_t}^k,$$

其预对偶为 $W_1 \times^{\|\cdot\|_\infty} \cdots \times W_t$。
求值函数 $\delta_\theta\colon W_i \to \mathbb{R}$ 是连续的，对每个 $\theta \in \Theta$均成立。事实上，对每个$\mathbf{x} \in \mathbb{R}^s$,

$$\left|\delta_\theta\left(\sum_{j=1}^k \alpha_j k_i(\mathbf{x},\cdot)\right)\right| = \left|\sum_{j=1}^k \alpha_j k_i(\mathbf{x},\theta)\right| \le \left\|\sum_{j=1}^k \alpha_j k_i(\mathbf{x},\cdot)\right\|_\infty .$$

于是，根据定理 6.9，另见例6.12,

$$\overline{\text{Lin}\{k(\cdot,\theta) : \theta\in\Theta\}}^{w^*} = \overline{\text{Lin}\{f_\theta^{\text{net}} : \theta\in\Theta\}}^{w^*} = B_{\Phi^0} = B_\Phi.$$

近期研究（Bohn、Griebel 和 Rieger，2018年；Wang、Xu 和 Yan，2025年；Heeringa、Spek 和 Brune，2025年）表明，固定架构的深度神经网络训练可以严格地表述为再生核巴拿赫空间（RKBS）中的最小范数问题，这与本文采用的方法一致。在这一设定下，核

$$k : X \times X' \to \mathbb{R},$$

由输入域 $X = \mathbb{R}^s$ 和参数空间 $X' = \Theta$ 共同决定。因此，对于固定架构和参数 $\theta \in \Theta$ 的神经网络，给定一个训练数据集

$$\mathcal{D}_m := \{(\mathbf{x}_j, \boldsymbol{y}_j)\}_{j=1}^m \subset \mathbb{R}^s \times \mathbb{R}^t,$$

等价于在再生核巴拿赫空间 B_Φ 中 求解最小范数插值问题，其中 b y

$$\arg\min\{\|f_\mu\| \;:\; f_\mu \in \mathcal{M}_{\mathbf{y}}\}, \tag{7.7}$$

其中 $\mathcal{M}_{\boldsymbol{y}} := \{f_\mu \in B_\Phi\colon\; f_\mu(\mathbf{x}_j) = \boldsymbol{y}_j,\;\text{且}\;\; j = 1,\;\cdots,\;m\}.$
这种方法是合理的，因为神经网络可以自然地解释为一种特殊的向量值再生核巴拿赫空间中的元素。在这种情况下， B_Φ 表示包含每个神经网络 f_θ^{net} 的 最小巴拿赫空间，适用于每个 $\theta \in \Theta$, 并且在 $\mathbf{x}$ 上关于逐点收敛是闭合的，正如 $\text{Lin}\{f_\theta^{\text{net}}\colon\; \theta\in\Theta\}^{w^*} = B_\Phi$.
由于 $B_\Phi = B_\Phi^k := B_{\Phi_1}^k \times^{\|\cdot\|_1} \cdots \times B_{\Phi_t}^k$，对于每个 $f_\mu \in B_\Phi = B_\Phi^k$，我们有

$$\|f_\mu\| := \sum_{i=1}^t \|\pi_i f_\mu\|,$$

并且显然 $\|f_\mu\|$ 在 $\|\pi_i f_\mu\|$ 对每个 $i = 1$、$\cdots$、t. 因此，训练问题（7.7）可分解为 t 个标量最小范数插值问题。更确切地说，令

$$\mathcal{M}_{\boldsymbol{y}}^i := \{\,\pi_i f_\mu : \mathbb{R}^s \to \mathbb{R} \;:\; g(\mathbf{x}_j) = \pi_i(y_j),\;\; \text{with } j = 1, \ldots, m\,\},$$

（7.7）的解为

$$\arg\min\left\{\|\pi_i f_\mu\| \;:\; \pi_i f_\mu \in \mathcal{M}_{\boldsymbol{y}}^i\right\}, \quad \text{for every } \;\; i = 1, \ldots, t.$$

如定理4.1所证明的，这些问题的解由相应数据函数的哈恩-巴拿赫延拓 $\widehat{\mu_i}^{\mathrm{data}}$导出 ${\mu_i}^{\mathrm{data}}$。实际上，这些解是函数 $f_{\widehat{\mu_i}^{\mathrm{data}}}$，对于每个 $i=1\cdots t$。

总之， $B_\Phi := B_{\Phi_1} \times \overset{\|\cdot\|_1}{\cdots} \times B_{\Phi_t}$，训练过程按分量进行。除了定理4.1之外，表示定理（定理 5.1）显然也自然适用，连同围绕它建立的所有结果。更准确地说，我们可以利用推论 5.4 和推论 5.6。基于这些考虑，接下来我们给出一种处理方法。

如前所述，将我们的分析限制在 $t=1$ 的情形不会失去一般性。然后，按如下步骤进行：

（i）从给定数据集中获取最大良好数据集

$$\mathcal{D}_m := \{(\mathbf{x}_j, \boldsymbol{y}_j)\}_{j=1}^m \subset \mathbb{R}^s \times \mathbb{R}.$$

(ii) 找到 $\theta_1, \cdots, \theta_m$，使得 $M = (k(\mathbf{x}_i, \theta_j))_{ij}$ 是对角的，其中 $|k(\mathbf{x}_i, \theta_i)| = \|k(\cdot, \theta_i)\|$.

(iii) 确定 $\mathbb{R}^m$ 中相应的可容许符号向量集合。

(iv) 如果 $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m)$ 和 $\mathbf{y} = (y_1, \ldots, y_m)$ 满足 $M\boldsymbol{\beta}^T = \boldsymbol{y}^T$ 且 $\operatorname{sign}(\boldsymbol{\beta})$ 是一个可容许符号向量，那么 $f_\mu(\cdot) = \sum_{i=1}^m \beta_i k(\cdot, \theta_i)$ 是方程 (7.7) 的一个解。

(v) 如果 $\operatorname{sign}(\beta)$ 不可容许，那么我们可以处理以下对象的正则化：

$$f_\mu(\cdot) = \sum_{i=1}^m \beta_i k(\cdot, \theta_i).$$

为此，考虑所有具有 $\boldsymbol{\beta}_{reg} = (\beta_1, \cdots, \beta_m)$ 这种形式且其符号向量 $\operatorname{sign}(\beta_{reg})$，可容许的向量。如果 $\boldsymbol{\beta}_{reg} = (\beta_1, \cdots, \beta_m)$ 是这样的一个向量，那么我们称 $f_\mu(\cdot) = \sum_{i=1}^m \beta_i k(\cdot, \theta_i)$ 是 f_μ. 的一个正则化。注意，$f_\mu(\cdot)$ 是该问题的一个解

$$\arg\min\{\|f_\mu\| : f_\mu \in \mathcal{M}_{\tilde{\mathbf{y}}}\},$$

其中 $M\boldsymbol{\beta}_{reg}^T = \tilde{\mathbf{y}}^T$。然后，根据正则化问题中定义的函数 $Q_{\mathbf{y}}$ 和参数 λ_0，在 $f_\mu(\cdot) = \sum_{i=1}^m \beta_i k(\cdot, \theta_i)$ 与所有正则化 $f_\mu(\cdot) = \sum_{i=1}^m \beta_i k(\cdot, \theta_i)$，之间选择最佳选项

$$\inf\left\{Q_{\mathbf{y}}(\mathcal{L}(f_\mu)) + \lambda_0\, \varphi(\|f_\mu\|) : f_\mu \in B_\Phi\right\}.$$

回顾一下，对于神经网络 $f_{\mathrm{net}\theta}: \mathbb{R}^s \to \mathbb{R}^t$，该问题可分解为 t标量最小范数插值问题（见第 4节），只需考虑 $\pi_i f_\theta^{\mathrm{net}}: \mathbb{R}^s \to \mathbb{R}$，对于 $i=1\cdots, t.$

在下一节中，我们将给出一个示例，说明如何通过数据集学习函数这一过程。我们还将验证，为了处理良好数据集，获得对角矩阵 $M = (k(\mathbf{x}_i, \theta_j))_{ij}$ 其中

$$k(\mathbf{x}_i, \theta_i) = \|k(\cdot, \theta_i)\|,$$

在许多情况下，这是相当直接的。

8 说明性示例：作为特征再生核巴拿赫空间的神经网络

本节给出一个具体的说明性示例，展示如何在简单情形下实例化前几节建立的理论框架。本例的目的并非模拟真实的大规模学习问题，而是明确说明特征再生核巴拿赫空间如何从固定的神经网络架构中产生，以及相关的学习表述和表示定理结果如何应用于实践。

为便于说明，我们考虑一个小型神经网络架构，其中包含两个输入节点、一个由两个神经元组成的隐藏层，以及一个含三个节点的输出层。该架构足够简单，可以将所有构造明确写出，同时仍能体现理论的核心特征，包括诱导的特征映射、在所实现函数空间上相应的巴拿赫范数，以及由此得到的核表示。通过这一示例，我们说明，在诱导的特征再生核巴拿赫空间中进行最小范数插值，会得到符合前文所建立的表示定理的有限表示。

此类网络的结构为 $f_\theta^{\mathrm{net}}: \mathbb{R}^2 \to \mathbb{R}^3$。对于每个 $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$，我们有

$$f_\theta^{\mathrm{net}}(\boldsymbol{x}) = (f_\theta^{\mathrm{net},1}(\mathbf{x}), f_\theta^{\mathrm{net},2}(\mathbf{x}), f_\theta^{\mathrm{net},3}(\mathbf{x})) \in \mathbb{R}^3,$$

其中

$$\mathbf{x}=\begin{pmatrix}x_1\\x_2\end{pmatrix}\longmapsto f_\theta^{\text{net}}(\mathbf{x})=\begin{pmatrix}f_\theta^{\text{net},1}(\mathbf{x})\\f_\theta^{\text{net},2}(\mathbf{x})\\f_\theta^{\text{net},3}(\mathbf{x})\end{pmatrix},$$

$$\begin{aligned}f_\theta^{\text{net},1}(\mathbf{x})&=w_5\sigma(w_1x_1+w_2x_2+b_1)+w_6\sigma(w_3x_1+w_4x_2+b_2)+b_3,\\f_\theta^{\text{net},2}(\mathbf{x})&=\widetilde{w}_5\sigma(\widetilde{w}_1x_1+\widetilde{w}_2x_2+\widetilde{b}_1)+\widetilde{w}_6\sigma(\widetilde{w}_3x_1+\widetilde{w}_4x_2+\widetilde{b}_2)+\widetilde{b}_3,\\f_\theta^{\text{net},3}(\mathbf{x})&=\widehat{w}_5\sigma(\widehat{w}_1x_1+\widehat{w}_2x_2+\widehat{b}_1)+\widehat{w}_6\sigma(\widehat{w}_3x_1+\widehat{w}_4x_2+\widehat{b}_2)+\widehat{b}_3.\end{aligned}$$

此处，$\sigma(t)=\max\{t,0\}$表示每一个 $t\in\mathbb{R}$的 ReLU 激活函数。我们将考虑训练数据集

$$\mathcal{D}_3=\{(\mathbf{x}_1,\boldsymbol{y}_1),(\mathbf{x}_1,\boldsymbol{y}_2),(\mathbf{x}_3,\boldsymbol{y}_3)\}$$

其中，输入数据$\mathbf{x}_1=(1,-1),\mathbf{x}_2=(-1,0),\mathbf{x}_3=(0,1)$属于 $\mathbb{R}^2$,输出数据 $\boldsymbol{y}_1=(a_1,b_1,c_1),\boldsymbol{y}_2=(\widetilde{a},\widetilde{b},\widetilde{c})$, 以及 $\boldsymbol{y}_3=(\widehat{a},\widehat{b},\widehat{c})$属于 $\mathbb{R}^3$.

我们已在第7 节中证明，训练整个神经网络 f_θ^{net}等价于分别训练其各个输出分量；$f^{\text{net}}{}_{,\theta}{}^{1}(\mathbf{x})$ $f_{\text{net}_\theta}{}^{2}(\mathbf{x})$,以及 $f_{\text{net}_\theta}{}^{3}(\mathbf{x})$。因此，我们将注意力限制于一个网络的训练

$$f_\theta^{\text{net}}:\mathbb{R}^2\to\mathbb{R},$$

由 $f_\theta^{\text{net}}(\mathbf{x})=w_5\,\sigma(w_1x_1+w_2x_2+b_1)+w_6\,\sigma(w_3x_1+w_4x_2+b_2)+b_3$，其中 $\sigma(t)=\max\{t,0\}$ 表示 ReLU 激活函数。因此，我们将考虑以下训练数据集

$$\mathcal{D}_3:=\{(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),(\mathbf{x}_3,y_3)\},$$

其中 $\mathbf{x}_1=(1,\ -1)$、$\mathbf{x}_2=(-1,\ 0)$、$\mathbf{x}_3=(0,\ 1)$以及 $y_1=a$、$y_2=b$、$y_3=c$.

在此框架中,

$$\|\theta\|:=\max\{|w_1|+|w_2|+|b_1|\,;|w_3|+|w_4|+|b_2|\,;|w_5|+|w_6|+|b_3|\}\qquad(\theta\in\Theta).$$

其中

$$\theta=\left(\begin{pmatrix}w_1&w_2\\w_3&w_4\end{pmatrix},\begin{pmatrix}b_1\\b_2\end{pmatrix},(w_5,w_6),b_3\right)=(w_1,w_2,w_3,w_4,b_1,b_2,w_5,w_6,b_3)\in\Theta.$$

由于 Θ 是有限维的，其上的所有范数都等价。尽管如此，我们仍可选择任意方便的范数。

为了确定要使用的核函数，我们将 $\xi:\Theta\to\mathbb{R}^+$ 定义为

$$\xi(\theta):=\frac{1}{\max\{1,\|\theta\|^3\}},\quad\theta\in\Theta.\tag{8.1}$$

现在，我们将核函数 $k:\ \mathbb{R}^2\times\Theta\to\mathbb{R}$ 定义为

$$k(\mathbf{x},\theta):=f_\theta^{\text{net}}(\mathbf{x})\,\xi(\theta),\ \text{ for every }\mathbf{x}=(x_1,x_2)\in\mathbb{R}^2\text{ and }\theta\in\Theta.$$

注意，对每个 $\mathbf{x}\in\mathbb{R}^2$，都有 $k(\mathbf{x},\cdot)\in\mathcal{B}(\Theta)$，其中 $\mathcal{B}(\Theta)$ 表示 Θ上所有连续有界函数构成的巴拿赫空间，并赋予上确界范数 $\|\cdot\|_\infty$。事实上，

$$\lim_{\|\theta\|\to\infty}f_\theta^{\text{net}}(\mathbf{x})\,\xi(\theta)=0,$$

由此可知，对每个 $\mathbf{x}\in\mathbb{R}^2$，都有 $k(\mathbf{x},\cdot)\in C_0(\Theta)$。

The因此，对于 $\theta=(w_1,w_2,w_3,w_4,b_1,b_2,w_5,w_6,b_3)\in\Theta$ 以及 $\mathbf{x}=(x_1,x_2)\in\mathbb{R}^2$，我们有

$$k(\mathbf{x},\theta)=\frac{w_5\,\sigma(w_1x_1+w_2x_2+b_1)+w_6\,\sigma(w_3x_1+w_4x_2+b_2)+b_3}{\max\{1,\|\theta\|^3\}},$$

因此,

$$\begin{aligned}|k(\mathbf{x},\theta)|&\le\frac{|w_5|\big(|w_1x_1|+|w_2x_2|+|b_1|\big)+|w_6|\big(|w_3x_1|+|w_4x_2|+|b_2|\big)+|b_3|}{\max\{1,\|\theta\|^3\}}\\&\le\frac{\max\{|x_1|,|x_2|,1\}\Big(\|\theta\|\,|w_5|+\|\theta\|\,|w_6|+|b_3|\Big)}{\max\{1,\|\theta\|^3\}}\\&\le\frac{\max\{|x_1|,|x_2|,1\}\|\theta\|^2}{\max\{1,\|\theta\|^3\}}\le\max\{|x_1|,|x_2|,1\}.\end{aligned}\tag{8.2}$$

另一方面，若 $\max\{|x_1|,|x_2|,1\}=|x_1|$，则对于 $\theta_0=(\pm 1,0,0,0,0,0,1,0,0)$（我们根据 x_1的符号选择 $+1$ 或 -1），我们有

$$k(\mathbf{x},\theta_0)=\sigma(|x_1|)=|x_1|.$$

类似地，如果 $\max\{|x_1|,|x_2|,1\}=|x_2|$，那么对于 $\theta_0=(0,\ \pm 1,\ 0,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0)$，我们有 $k(\mathbf{x}\ \theta_0)=|x_2|$。此外，如果 $\max\{|x_1|,|x_2|,1\}=1$，那么对于 $\theta_0=(0,\ 0,\ 0,\ 0,\ 1,\ 0,\ 1,\ 0,\ 0)$，我们有 $k(\mathbf{x}\ \theta_0)=\sigma(1)=1$。因此，

$$\|k(\mathbf{x},\cdot)\|_\infty=\max\{|x_1|,|x_2|,1\}.$$

回顾一下，我们的数据集为 $\{(\mathbf{x}_1,\ y_1),\ (\mathbf{x}_2,\ y_2),\ (\mathbf{x}_3,\ y_3)\}$，其中

$$\mathbf{x}_1=(1,-1),\qquad \mathbf{x}_2=(-1,0),\qquad \mathbf{x}_3=(0,1),$$

以及 $\boldsymbol{y}=(y_1,\ y_2,\ y_3)=(a,\ b,\ c)$。因此，

$$\|k(\mathbf{x}_i,\cdot)\|_\infty=1,\qquad \text{for every } i=1,2,3.$$

此外，
(i) 对于 $\theta_1:=(1,0,0,0,0,0,0,1,0,0)$，我们有 $k(\mathbf{x}\ \theta_1):=\sigma(x_1)$，因此 $k(\mathbf{x}1,\theta_1)=1$ 且 $k(\mathbf{x}2,\theta_1)=k(\mathbf{x}3,\theta_1)=0$.
(ii) 对于 $\theta_2:=(-1,0,0,0,0,0,0,1,0,0)$，我们有 $k(\mathbf{x}\ \theta_2):=\sigma(-x_1)$，因此 $k(\mathbf{x}2,\theta_2)=1$，而 $k(\mathbf{x}1,\theta_2)$ 和$k(\mathbf{x}3,\theta_2)$ 均等于 0。
(iii) 对于 $\theta_3:=(0,1,0,0,0,0,0,1,0,0)$，我们有 $k(\mathbf{x}\ \theta_3):=\sigma(x_2)$，因此 $k(\mathbf{x}_3,\theta_3)=1$，而 $k(\mathbf{x}_1,\theta_3)$ 和$k(\mathbf{x}_2,\theta_3)$ 均等于 0。

因此，

$$k(\mathbf{x}_i,\theta_j)=0,\quad \text{if}\quad i\neq j,\ \text{and}\ k(\mathbf{x}_i,\theta_i)=\|k(\cdot,\theta_i)\|=1,\quad \text{for}\quad i=1,2,3.$$

事实上，如果 $\sum_{i=1}^m \alpha_i k(\tilde{}\mathbf{x}_i,\cdot)\in \mathrm{Lin}\{k(\mathbf{x}\ \cdot):\mathbf{x}\in\mathbb{R}^2\}$ 满足 $\left\|\sum_{i=1}^m \alpha_i k(\tilde{}\mathbf{x}_i,\cdot)\right\|_\infty\le 1$，那么，$\left|\sum_{i=1}^m \alpha_i k(\tilde{}\mathbf{x}_i,\theta)\right|\le 1$，从而

$$\|k(\cdot,\theta)\|=\sup\left\{\left|\sum_{i=1}^m \alpha_i k(\widetilde{\mathbf{x}}_i,\theta)\right|:\left\|\sum_{i=1}^m \alpha_i k(\widetilde{\mathbf{x}}_i,\cdot)\right\|_\infty\le 1\right\}\le 1,$$

并且可以推出，如果对于某个 $\boldsymbol{x}\in\mathbb{R}^2$，$k(\mathbf{x}\ \theta)=1=\|k(\mathbf{x}\ \cdot)\|_\infty$，那么 $\|k(\cdot,\theta)\|=1$. 因此，

$$M:=\begin{pmatrix}k(\mathbf{x}_1,\theta_1) & k(\mathbf{x}_1,\theta_2) & k(\mathbf{x}_1,\theta_3)\\ k(\mathbf{x}_2,\theta_1) & k(\mathbf{x}_2,\theta_2) & k(\mathbf{x}_2,\theta_3)\\ k(\mathbf{x}_3,\theta_1) & k(\mathbf{x}_3,\theta_2) & k(\mathbf{x}_3,\theta_3)\end{pmatrix}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\tag{8.3}$$

现在来确定可容许符号向量的集合。符号空间为

$$\Delta:=\{\boldsymbol{s}=(s_1,s_2,s_3):s_i\in\{-1,0,1\},\ i=1,2,3\}.$$

根据定义 5.5，符号向量 $\boldsymbol{s}=(s_1,s_2,s_3)\in\Delta^3$ 可容许，当且仅当

$$|||\,\boldsymbol{s}\,|||:=\left\|\sum_{i=1}^3 s_i k(\mathbf{x}_i,\cdot)\right\|_\infty=\sup_{\theta\in\Theta}\left|\sum_{i=1}^3 s_i k(\mathbf{x}_i,\theta)\right|\le 1.$$

要确定可容许符号向量的集合，注意到 $\theta=(0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1)\in\Theta$ 满足 $\|\theta\|=1$ 且

$$\left|\sum_{i=1}^3 s_i k(\mathbf{x}_i,\theta)\right|=\left|\sum_{i=1}^3 s_i\right|,$$

因此，所有满足 $\left|\sum_{i=1}^3 s_i\right|>1$ 的符号向量 $\boldsymbol{s}=(s_1,\ s_2,\ s_3)$ 均不可容许。因此，向量 $\pm(1,\ 1,\ 0)$，$\pm(1,\ 0,\ 1)$ $\pm(0,\ 1,\ 1)$ $\pm(1,\ 1,\ 1)$ 均不可容许。此外：
(i) 对于 $\theta:=(-\frac{4}{15},\frac{2}{15},\ 0,\ 0,\frac{1}{3},\ 0,\ 1,\ 0,\ 0)$，我们有

$$|-k(\mathbf{x}_1,\theta)+k(\mathbf{x}_2,\theta)+k(\mathbf{x}_3,\theta)|>1,$$

因此 $\pm(-1,\ 1,\ 1)$ 不可容许。

(ii) 对于 $\theta:=(\frac{3}{10},\ \frac{1}{5},\ 0,\ 0,\ \frac{1}{2},\ 0,\ 1,\ 0,\ 0)$，我们有

$$|k(\mathbf{x}_1,\theta)-k(\mathbf{x}_2,\theta)+k(\mathbf{x}_3,\theta)|>1,$$

因此，$\pm(1,\ -1,\ 1)$ 不可容许。

(iii) 对于 $\theta:=(\frac{1}{5},\ -\frac{3}{10},\ 0,\ 0,\ \frac{1}{2},\ 0,\ 1,\ 0,\ 0)$，我们有

$$|k(\mathbf{x}_1,\theta)+k(\mathbf{x}_2,\theta)-k(\mathbf{x}_3,\theta)|>1,$$

因此，$\pm(1,\ 1,\ -1)$ 不可容许。

另一方面，我们断言 $\pm(1,\ -1,\ 0)$ $\pm(1,\ 0,\ -1)$ $\pm(0,\ 1,\ -1)$ 是可容许的。事实上，对于 $\mathbf{x}=(x_1,\ x_2)$，$\widetilde{\mathbf{x}}=(\widetilde{x}_1,\ \widetilde{x}_2)$，以及 $\theta\in\Theta$，考虑

$$k(\mathbf{x},\theta)=\frac{w_5\sigma(a)+w_6\sigma(c)+b_3}{\max\{1,||\theta||^3\}}\ \text{ and }\ k(\widetilde{\mathbf{x}},\theta)=\frac{w_5\sigma(b)+w_6\sigma(d)+b_3}{\max\{1,||\theta||^3\}},$$

其中辅助变量 $a,\ b,\ c,\ d$ 定义为

$$\begin{aligned} a&=w_1x_1+w_2x_2+b_1, & b&=w_1\widetilde{x}_1+w_2\widetilde{x}_2+b_1,\\ c&=w_3x_1+w_4x_2+b_2, & d&=w_3\widetilde{x}_1+w_4\widetilde{x}_2+b_2. \end{aligned}$$

于是，

$$\begin{aligned} |k(\mathbf{x},\theta)-k(\widetilde{\mathbf{x}},\theta)|&=\frac{\big|\,w_5\big(\sigma(a)-\sigma(b)\big)+w_6\big(\sigma(c)-\sigma(d)\big)\,\big|}{\max\{1,\|\theta\|^3\}}\\ &\le\frac{|w_5|\,|\sigma(a)-\sigma(b)|+|w_6|\,|\sigma(c)-\sigma(d)|}{\max\{1,\|\theta\|^3\}}. \end{aligned}$$

由于，$|\sigma(a)-\sigma(b)|\le\max\{|\sigma(a)|,\ |\sigma(b)|\}$ 和 $|\sigma(c)-\sigma(d)|\le\max\{|\sigma(c)|,\ |\sigma(d)|\}$，我们有

$$\begin{aligned} |k(\mathbf{x},\theta)-k(\tilde{\mathbf{x}},\theta)|&\le\frac{\max\{|\sigma(a)|,|\sigma(b)|,|\sigma(c)|,|\sigma(d)|\}\,(|w_5|+|w_6|)}{\max\{1,\|\theta\|^3\}}\\ &\le\frac{\max\{|x_1|,|x_2|,|\tilde{x}_1|,|\tilde{x}_2|,1\}||\theta||^2}{\max\{1,\|\theta\|^3\}}\le\max\{|x_1|,|x_2|,|\tilde{x}_1|,|\tilde{x}_2|,1\}. \end{aligned}$$

由此可知，$\pm(1,\ -1,\ 0)$ $\pm(1,\ 0,\ -1)$ $\pm(0,\ 1,\ -1)$ 是可容许的。此外，向量 $\pm(1,\ 0,\ 0)$ $\pm(0,\ 1,\ 0)$，以及$\pm(0,\ 0,\ 1)$ 是可容许的，因为 $||k(\mathbf{x}_i,\ \cdot)||_\infty=1$，对于 $i=1,\ 2,\ 3$。因此，我们得出可容许符号向量的集合为

$$\Delta^3{}_{\mathrm{adm}}=\{(0,0,0),\pm(1,0,0),\pm(0,1,0),\pm(0,0,1),\pm(1,-1,0),\pm(1,0,-1),\pm(0,1,-1)\}.$$

为了训练我们的网络，由于 $M=(k(x_i,\ \theta_j))_{ij}=I_3$，显然有 $\boldsymbol{\beta}=(a,\ b,\ c)=\boldsymbol{y}$. 若 $\mathrm{sign}(\boldsymbol{y})$ 可容许，则

$$f^{\mathrm{sol}}(\cdot)=a\,k(\cdot,\theta_1)+b\,k(\cdot,\theta_2)+c\,k(\cdot,\theta_3) \tag{8.4}$$

是以下问题的解

$$\arg\min\{\|f_\mu\|:f_\mu\in\mathcal{M}_{\boldsymbol{y}}\}. \tag{8.5}$$

事实上，$\|f^{\mathrm{sol}}\|=|a|+|b|+|c|$ 且该值在 $s_1k(x_1,\ \cdot)+s_2k(x_2,\ \cdot)+s_3k(x_3,\ \cdot)$ 处取得。

例如，若 $a<0,\ b=0$ 且 $c>0$，则 $f^{\mathrm{sol}}=ak(x_1,\ \cdot)+ck(x_3,\ \cdot)$ 是 (8.5) 的解，其中 $\mathbf{y}=(a,\ 0,\ c)$，因为$(-1,\ 0,\ 1)$ 是一个可容许符号向量。

现在假设，例如，$\mathbf{y}=(a,\ b,\ c)=(2,\ -3,\ \frac{1}{2})$。那么，$\mathrm{sign}(\boldsymbol{y})$ 不可容许。因此，我们考虑如下类型的正则化问题

$$\arg\min\big\{Q_{\mathbf{y}}(\mathcal{L}(f_\mu))+\lambda_0\,\varphi(\|f_\mu\|):f_\mu\in B_\Phi\big\}. \tag{8.6}$$

举例来说，令

$$\mathcal{R}(f_\mu):=Q_{\mathbf{y}}(\mathcal{L}(f_\mu))+\lambda_0\,\varphi(\|f_\mu\|)=\frac{(\tilde{y}_1-a)^2+(\tilde{y}_2-b)^2+(\tilde{y}_3-c)^2}{3}+\lambda_0\|f_\mu\|,$$

其中 $\mathcal{L}(f_\mu) = (f_\mu(x_1)\ f_\mu(x_2)\ f_\mu(x_3)) = (\tilde{y}_1,\ \tilde{y}_2,\ \tilde{y}_3)$ 且 $\lambda_0 = \frac{1}{10^\circ}$

在这种情况下，我们有映射 $\widehat{\mu}^{\text{data}}\colon W \to \mathbb{R}$，其定义为

$$\widehat{\mu}^{\text{data}} = 2\delta_{\theta_1} - 3\delta_{\theta_2} + \tfrac{1}{2}\delta_{\theta_3}$$

是由等式确定的线性映射 $\mu^{\text{data}}\colon W^{\text{data}} \to \mathbb{R}$ 的延拓

$$\mu^{\text{data}}(k(x_1,\cdot)) = 2, \qquad \mu^{\text{data}}(k(x_2,\cdot)) = -3, \qquad \mu^{\text{data}}(k(x_3,\cdot)) = \tfrac{1}{2}.$$

此外，

$$f_{\widehat{\mu}^{\text{data}}} = 2k(\cdot,\theta_1) - 3k(\cdot,\theta_2) + \tfrac{1}{2}k(\cdot,\theta_3).$$

然而，我们不知道 $\widehat{\mu}^{\text{data}}$ 是否是 μ^{data} 的哈恩–巴拿赫延拓。因此，我们还考虑如下形式的正则化解

$$f_\varepsilon(\cdot) = (2+\varepsilon_1)k(\cdot,\theta_1) + (-3+\varepsilon_2)k(\cdot,\theta_2) + \big(\tfrac{1}{2}+\varepsilon_3\big)k(\cdot,\theta_3),$$

其中 $\boldsymbol{\beta}_\varepsilon = (2+\varepsilon_1,\ -3+\varepsilon_2, -\frac{1}{2}+\varepsilon_3)$，且 $\operatorname{sign}(\boldsymbol{\beta}_\varepsilon)$ 是可容许的。这些就是

$$\arg\min\{\|f_\mu\| : f_\mu \in \mathcal{M}_{\mathbf{y}_\varepsilon}\},$$

其中 $\mathbf{y}_\varepsilon^T = M\boldsymbol{\beta}_\varepsilon^T = (2+\varepsilon_1,\ -3+\varepsilon_2,\ \frac{1}{2}+\varepsilon_3)^T$。对于每个可容许符号向量，我们选择最优的 $\varepsilon = (\varepsilon_1,\ \varepsilon_2,\ \varepsilon_3)$，即使 (8.6) 取最小值的那个。

首先，注意到 $f_0(\cdot) = 2k(\cdot,\theta_1) - 3k(\cdot,\theta_2) + \frac{1}{2}k(\cdot,\theta_3)$ 满足 $\mathcal{R}(f_0) = \lambda_0\|f_0\|$，其中 $\|f_0\| \le 2+3+\frac{1}{2} = \frac{11}{2}$。此外，由于 $|k(x_1,\cdot) - k(x_2,\cdot)\| \le 1$，因此 $\|f_0\| \ge |2(k(x_1,\theta_1) + 3k(x_2,\theta_3))| = 5$。因此，$0.5 \le \mathcal{R}(f_0) \le 0.55$.现在，我们确定与每个可容许符号向量相关联的最优正则化解。

(i) 对于 $\mathbf{s}_1 := (1,-1,0)$，f_0 的正则化形式为

$$f_{\boldsymbol{s_1}}(\cdot) = (2+\varepsilon_1)k(\cdot,\theta_1) + (-3+\varepsilon_2)k(\cdot,\theta_2),$$

其中 $2+\varepsilon_1 \ge 0$ 和 $-3+\varepsilon_2 \le 0$。因此，

$$\mathcal{R}(f_{\boldsymbol{s_1}}) = \frac{\varepsilon_1^2+\varepsilon_2^2+\frac{1}{4}}{3} + \frac{1}{10}(|2+\varepsilon_1| + |-3+\varepsilon_2|) = \frac{\varepsilon_1^2+\varepsilon_2^2}{3} + \frac{1}{12} + \frac{1}{10}(\varepsilon_1-\varepsilon_2) + \frac{1}{2}.$$

$\mathcal{R}(f_{\boldsymbol{s_1}})$ 的最小值在 $(\varepsilon_1,\varepsilon_2) = \big(-\frac{3}{20}, \frac{3}{20}\big)$处取得。因此，与符号向量 $\mathbf{s}_1 = (1,-1,0)$相关联时，我们对正则化解的最佳选择是

$$f_{\tilde{\mathbf{y}}_1}(\cdot) = \big(2-\tfrac{3}{20}\big)k(\cdot,\theta_1) + \big(-3+\tfrac{3}{20}\big)k(\cdot,\theta_2),$$

它是以下问题的一个解

$$\arg\min\{\|f_\mu\| : f_\mu \in \mathcal{M}_{\tilde{\mathbf{y}}_1}\},$$

其中 $\tilde{\mathbf{y}}_1 = \big(2-\frac{3}{20}, -3+\frac{3}{20}, 0\big)$。此外，

$$\|f_{\tilde{\mathbf{y}}_1}\| = |2-\tfrac{3}{20}| + |-3+\tfrac{3}{20}| = \tfrac{94}{20},$$

该范数在 $k(x_1,\cdot) - k(x_2,\cdot)$处取得，并且

$$\mathcal{R}(f_{\tilde{\mathbf{y}}_1}) = \frac{\frac{9}{400}+\frac{9}{400}+\frac{1}{4}}{3} + \frac{1}{10}\|f_{\tilde{\mathbf{y}}_1}\| = \frac{59}{600} + \frac{1}{10}\frac{94}{20} = \frac{341}{600} \approx 0.5683.$$

（ii）对于 $\mathbf{s}_2 := (-1,1,0)$，我们有

$$f_{\boldsymbol{s_2}}(\cdot) = (2+\varepsilon_1)k(\cdot,\theta_1) + (-3+\varepsilon_2)k(\cdot,\theta_2),$$

其中 $2+\varepsilon_1 \le 0$ 和 $-3+\varepsilon_2 \ge 0$。因此，我们最小化

$$\mathcal{R}(f_{\boldsymbol{s_2}}) = \frac{\varepsilon_1^2+\varepsilon_2^2+\frac{1}{4}}{3} + \frac{1}{10}(|2+\varepsilon_1| + |-3+\varepsilon_2|),$$

其中 $\varepsilon_1 \le -2$ 和 $\varepsilon_2 \ge 3$。最小值在 $(\varepsilon_1,\ \varepsilon_2) = (-2,\ 3)$ 处取得。因此，

$$f_{\tilde{\mathbf{y}}_2}(\cdot) = 0 \quad \text{with} \quad \mathcal{R}(f_{\tilde{\mathbf{y}}_2}) = \frac{\varepsilon_1^2+\varepsilon_2^2+\frac{1}{4}}{3} = \frac{4+9+\frac{1}{4}}{3} = \frac{53}{12}.$$

(iii) 对于 $\mathrm{s}_3 := (1,\ 0,\ -1)$ 和 $\mathrm{s}_4 := (-1,\ 0,\ 1)$，我们有

$$f_{\boldsymbol{s_j}}(\cdot) = (2+\varepsilon_1)k(\cdot,\theta_1) + \big(\tfrac{1}{2}+\varepsilon_3\big)k(\cdot,\theta_3),$$

（其中 $j=3,\ 4$），其中 $2+\varepsilon_1 \geq 0$ 且 $\frac{1}{2}+\varepsilon_3 \leq 0$ 若 $j=3$，并且 $2+\varepsilon_1 \leq 0$ 且 $\frac{1}{2}+\varepsilon_3 \geq 0$ 若 $j=4$。其最小值

$$\mathcal{R}(f_{\boldsymbol{s_j}}) = \frac{\varepsilon_1^2+9+\varepsilon_3^2}{3} + \frac{1}{10}\Big(|2+\varepsilon_1| + \big|\tfrac{1}{2}+\varepsilon_3\big|\Big),$$

对于 $j=3$，其在 $f_{\tilde{\boldsymbol{y}}_3}(\cdot) = (2-\frac{3}{20})k(\cdot,\ \theta_1)$ 中取到，其中 $\mathcal{R}(f_{\tilde{\boldsymbol{y}}_3}) = \frac{3931}{1200} \approx 3.275$；与此同时，对于 $j=4$，其在 $f_{\tilde{\boldsymbol{y}}_4}(\cdot) = \big(\frac{1}{2}-\frac{3}{20}\big)k(\cdot,\ \theta_3)$ 中取到，其中 $\mathcal{R}(f_{\tilde{\boldsymbol{y}}_4}) = \frac{5251}{1200} \approx 4.375$。

(iv) 对于 $\mathrm{s}_5 := (0,\ -1,\ 1)$ 和 $\mathrm{s}_6 := (0,\ 1,\ -1)$，我们有

$$f_{\boldsymbol{s_5}}(\cdot) = f_{\boldsymbol{s_6}}(\cdot) = (-3+\varepsilon_2)k(\cdot,\theta_2) + \big(\tfrac{1}{2}+\varepsilon_3\big)k(\cdot,\theta_3),$$

其中 $-3+\varepsilon_2 \leq 0$ 且 $\frac{1}{2}+\varepsilon_3 \geq 0$ 对于 $f_{\boldsymbol{s_5}}(\cdot)$，以及 $-3+\varepsilon_2 \geq 0$ 对于 $\frac{1}{2}+\varepsilon_3 \leq 0$（$f_{\boldsymbol{s_6}}$）。注意，对于 $j=5,\ 6$,

$$\mathcal{R}(f_{\boldsymbol{s_j}}) = \frac{4+\varepsilon_2^2+\varepsilon_3^2}{3} + \frac{1}{10}\Big(|-3+\varepsilon_2| + \big|\tfrac{1}{2}+\varepsilon_3\big|\Big).$$

我们得到的最小值为

$$f_{\tilde{\boldsymbol{y}}_5}(\cdot) = (-3+\tfrac{3}{20})k(\cdot,\theta_2) + \big(\tfrac{1}{2}-\tfrac{3}{20}\big)k(\cdot,\theta_3) \text{ with } \mathcal{R}(f_{\tilde{\boldsymbol{y}}_5}) = \tfrac{1001}{600} \approx 1.66,$$

以及 $f_{\tilde{\boldsymbol{y}}_6}(\cdot) = 0$ 与 $\mathcal{R}(f_{\boldsymbol{\varepsilon}_6}) = \frac{53}{12} \approx 4.411.$

(v) 对于 $\mathrm{s}_7 := (1,\ 0,\ 0)$ 和 $\mathrm{s}_8 := (-1,\ 0,\ 0)$，我们有

$$f_{\boldsymbol{s_j}}(\cdot) = (2+\varepsilon_1)k(\cdot,\theta_1),$$

对于 $j=7,\ 8$，其中 $2+\varepsilon_1 \geq 0$ 若 $j=7$，且 $2+\varepsilon_1 \leq 0$ 若 $j=8$。由此得到

$$\mathcal{R}(f_{\boldsymbol{s_j}}) = \frac{\varepsilon_1^2+9+\frac{1}{4}}{3} + \frac{1}{10}|2+\varepsilon_1|, \qquad (j=7,8).$$

因此,

$$\mathcal{R}(f_{\boldsymbol{s_7}}) = \frac{\varepsilon_1^2+9+\frac{1}{4}}{3} + \frac{1}{10}(2+\varepsilon_1),$$

其中 $2+\varepsilon_1 \geq 0$，且

$$\mathcal{R}(f_{\boldsymbol{s_8}}) = \frac{\varepsilon_1^2+9+\frac{1}{4}}{3} - \frac{1}{10}(2+\varepsilon_1),$$

其中 $2+\varepsilon_1 \leq 0$。我们得到

$$f_{\tilde{\mathbf{y}}_7}(\cdot) = (2-\tfrac{3}{20})k(\cdot,\theta_1), \quad \mathcal{R}(f_{\tilde{\mathbf{y}}_7}) = \tfrac{3921}{1200} \approx 3.26, \quad f_{\tilde{\mathbf{y}}_8} = 0, \quad \mathcal{R}(f_{\tilde{\mathbf{y}}_8}) = \tfrac{53}{12} \approx 4.41.$$

(vii) 对于 $\mathrm{s}_9 := (0,1,0)$ 与 $\mathrm{s}_{10} := (0,-1,0)$，我们有

$$f_{\boldsymbol{s_j}}(\cdot) = (-3+\varepsilon_2)k(\cdot,\theta_1), \quad (j=9,10),$$

其中当 $j=9$ 时取 $-3+\varepsilon_2 \geq 0$，且当 $j=10$ 时取 $-3+\varepsilon_2 \leq 0$。因此,

$$\mathcal{R}(f_{\boldsymbol{s_j}}) = \frac{4+\varepsilon_2^2+\frac{1}{4}}{3} + \frac{1}{10}|-3+\varepsilon_2| = \frac{\varepsilon_2^2}{3} + \frac{17}{12} + \frac{1}{10}|-3+\varepsilon_2|, \quad (j=9,10).$$

我们得到

$$f_{\tilde{\mathbf{y}}_9} = 0 \quad \text{with} \quad \mathcal{R}(f_{\tilde{\mathbf{y}}_9}) = \tfrac{53}{12} \approx 4.41, \quad f_{\tilde{\mathbf{y}}_{10}}(\cdot) = (-3+\tfrac{3}{20})k(\cdot,\theta_1), \quad \mathcal{R}(f_{\tilde{\mathbf{y}}_{10}}) = \tfrac{2051}{1200} \approx 1.7.$$

(viii) 对 $\mathrm{s}_{11} := (0,\ 0,\ 1)$ 和 $\mathrm{s}_{12} := (0,\ 0,\ -1)$ 的情形类似，其中

$$f_{\boldsymbol{s_j}}(\cdot) = \big(\tfrac{1}{2}+\varepsilon_3\big)k(\cdot,\theta_3),$$

其中，$\frac{1}{2}\pm\varepsilon_3\geq 0$ 在 $j=11$ 时成立，$\frac{1}{2}+\varepsilon_3\leq 0$ 在 $j=12$ 时成立。于是，

$$\mathcal{R}(f_{\boldsymbol{\varepsilon}_j})=\frac{4+9+\varepsilon_3^2}{3}+\frac{1}{10}\Big|\tfrac{1}{2}+\varepsilon_3\Big|,\qquad (j=11,12).$$

因此，

$$f_{\tilde{\mathbf{y}}_{11}}(\cdot)=\big(\tfrac{1}{2}-\tfrac{3}{20}\big)k(\cdot,\theta_3),\quad \mathcal{R}(f_{\tilde{\mathbf{y}}_{11}})=\tfrac{5251}{1200}\approx 4.37,\quad f_{\tilde{\mathbf{y}}_{12}}(\cdot)=0,\quad \mathcal{R}(f_{\tilde{\mathbf{y}}_{12}})=\tfrac{53}{12}\approx 4.41.$$

(ix) 对于$\mathbf{s}_{13}:=(0,0,0)$，我们有 $f_{\tilde{\boldsymbol{y}}_{13}}(\cdot)=0$ 和 $\mathcal{R}(f_{\tilde{\boldsymbol{y}}_{13}})=\frac{53}{12}\approx 4.41$.

因此，在此框架下，我们的最佳选择是

$$f_0(\cdot):=f_{\widehat{\mu}^{\mathrm{data}}}=2k(\cdot,\theta_1)-3k(\cdot,\theta_2)+\tfrac{1}{2}k(\cdot,\theta_3).$$

注记 8.1. 如果 $(a,b,c)=\big(1,-\frac{1}{2},\frac{1}{2}\big)$ 和 $\lambda_0=\frac{1}{3}$，那么我们有

$$f_0:=k(\cdot,\theta_1)-\tfrac{1}{2}k(\cdot,\theta_2)+\tfrac{1}{2}k(\cdot,\theta_3),$$

取 $1+\frac{1}{2}\leq\|f_0\|\leq 2$，令 $f_0(k(\cdot,\ \theta_1)-\frac{1}{2}k(\cdot,\ \theta_2))=-\frac{3}{2}$。由此可得

$$\tfrac{1}{2}\leq\mathcal{R}(f_0)\leq\tfrac{2}{3}.$$

然而，对于 $\mathbf{s}_1:=(1,\ -1,\ 0)$，我们有，正则化项 Ω_0 为 f

$$f_{\boldsymbol{s}_1}(\cdot)=(1+\varepsilon_1)k(\cdot,\theta_1)+\big(-\tfrac{1}{2}+\varepsilon_2\big)k(\cdot,\theta_2),$$

取 $1+\varepsilon_1\geq 0$ 和 $-\frac{1}{2}+\varepsilon_2\leq 0$。因此，

$$\mathcal{R}(f_{\boldsymbol{s}_1})=\frac{\varepsilon_1^2+\varepsilon_2^2+\frac{1}{4}}{3}+\frac{1}{3}\big(|1+\varepsilon_1|+|-\tfrac{1}{2}+\varepsilon_2|\big)=\frac{\varepsilon_1^2+\varepsilon_2^2}{3}+\frac{1}{12}+\frac{1}{3}(\varepsilon_1-\varepsilon_2)+\frac{1}{2}.$$

最小值在 $f_{\tilde{\boldsymbol{y}}_1}(\cdot)=(1-{}_2)k(\cdot,\ \theta_1)$ 处取得，其中

$$\mathcal{R}(f_{\tilde{\boldsymbol{y}}_1})=\tfrac{1}{6}+\tfrac{1}{12}-\tfrac{1}{3}+\tfrac{1}{2}=\tfrac{5}{12}.$$

因此，在这种情况下，

$$f_{\tilde{\boldsymbol{y}}_1}(\cdot)=\tfrac{1}{2}k(\cdot,\theta_1)$$

比……更合适

$$f_0(\cdot)=k(\cdot,\theta_1)-\tfrac{1}{2}k(\cdot,\theta_2)+\tfrac{1}{2}k(\cdot,\theta_3).$$

注 8.2. 若在上述例子中，将 $\mathbf{x}_1:=(1,-1)$ 替换为 $\tilde{\ }\mathbf{x}_1:=(2,-2)$，那么对于相同的参数$\theta_1:=(1,0,0,0,0,0,1,0,0)$，我们有 $k(\mathbf{x}\ \theta_1):=\sigma(x_1)$，因此 $k(\tilde{\ }\mathbf{x}1,\theta_1)=2$，且 $k(\mathbf{x}2,\theta_1)=k(\mathbf{x}3,\theta_1)=0$。于是，$k(\cdot,\theta_1)$ 在 $\frac{1}{2}k(\tilde{\ }\mathbf{x}1,\cdot)$ 中达到其范数，我们按照上述方法处理 $\frac{1}{2}k(\tilde{\ }\mathbf{x}1,\cdot)$，从而得到一个解

$$f^{\mathrm{sol}}(\cdot)=\tfrac{a}{2}k(\cdot,\theta_1)+b\,k(\cdot,\theta_2)+c\,k(\cdot,\theta_3).$$

9 学习理论洞见与启示

本节提炼了本文所建立的特色再生核巴拿赫空间框架所带来的主要学习理论洞见。我们不再重述形式化构造，而是着重探讨这些结果如何改变对希尔伯特空间情形之外学习问题的理解，并阐明函数空间几何在现代学习模型中的作用。

第一个见解是，最小范数插值和正则化在巴拿赫空间中仍是概念上有意义的学习原则，但前提是施加额外的结构条件。与再生核希尔伯特空间不同，仅有点求值泛函的连续性，并不能确保特征表示、基于核的算法或有限维解结构的存在。从学习的角度来看，这解释了为何将基于再生核希尔伯特空间的方法简单地推广到非希尔伯特范数时往往会失败，也解释了为何由特征映射诱导的构造对于建立易于处理的学习问题表述至关重要。

第二个见解涉及表示定理的性质。在巴拿赫空间情形中，表示定理结果本质上是有条件的，即使插值问题适定，也可能不成立。这揭示了一个明确的概念边界：一边是具有核式解的学习问题，另一边则会产生真正的无限维解集。理解这条边界，对于阐释非二次正则化和诱导稀疏性的范数下学习算法的行为至关重要。

最后，将固定架构神经网络视为特殊向量值特征再生核巴拿赫空间中的元素，可提供统一的函数空间视角，将核方法、基于范数的正则化与神经网络架构联系起来。从这一视角看，架构选择和参数范数决定了所诱导的函数空间几何，进而决定学习模型的隐式归纳偏置。这一视角表明，神经网络学习的许多结构性方面，无需依赖无限宽极限或纯粹的希尔伯特空间假设，也可以得到理解。

10 结论

本研究构建了一个连贯的泛函分析框架，将经典的再生核希尔伯特空间理论拓展至巴拿赫空间情形，同时使之兼容现代学习模型。通过引入并系统分析特征再生核巴拿赫空间，我们明确了在希尔伯特空间情形之外，恢复基于核的学习表述和表示定理类结果所需的精确结构条件。

本分析的一项核心成果，是指出特征映射诱导结构和恰当的预对偶表示是巴拿赫空间中基于核的学习不可或缺的要素。在希尔伯特空间中，求值连续性已足够；但在巴拿赫空间情形下，通常还需要额外的几何与对偶假设。特征再生核巴拿赫空间框架将这些要求明确区分，并清晰地划分出支持核表示与不支持核表示的巴拿赫空间。

在这一框架下，我们证明了最小范数插值和正则化问题的存在性，并在明确条件下证明了表示定理。这些结果从形式上刻画了解何时具有有限核展开、何时只能获得弱-* 描述，从而阐明了经典核直觉在希尔伯特空间情形之外的适用边界。

我们进一步将该理论扩展到向量值特征再生核巴拿赫空间，为多输出学习问题奠定了严谨的理论基础。在此基础上，我们证明固定架构神经网络会自然诱导出特殊的向量值特征再生核巴拿赫空间，从而能够将网络训练严谨地表述为诱导函数巴拿赫空间中的一个变分问题。这建立了核方法、巴拿赫空间正则化与神经网络架构之间精确的数学联系。

本工作的研究范围有意聚焦于基础层面。我们不讨论优化动力学、统计泛化保证或计算效率，因为这些问题取决于额外的建模假设。相反，本文构建的框架为未来从统一的泛函分析视角研究隐式正则化、架构归纳偏置以及核-神经混合模型提供了结构基础。

未来研究的自然方向包括考察特征再生核巴拿赫空间中的优化动力学，尤其关注巴拿赫空间几何和特征映射结构如何影响隐式正则化以及基于梯度的算法的行为。另一个重要方向是在巴拿赫空间几何下推导统计泛化保证，包括纳入非希尔伯特范数和对偶结构影响的稳定性与鲁棒性分析。最后，开发能够利用特征再生核巴拿赫空间中特征映射诱导结构的高计算效率算法和逼近方案，仍是一项关键挑战；尤其是在大规模核方法和神经网络模型中，显式的函数空间表示并不可行。

致谢

本出版物是“伦理、负责任的通用人工智能：风险场景中的应用”（IAFER）项目的一部分，项目编号：TSI-100927-2023-1。该项目通过设立大学-产业研究计划（ENIA计划）获得资助，旨在开展人工智能的研究与开发，并在欧盟复苏、转型与韧性计划“欧盟下一代”的框架下，通过数字化转型与公务员事务部推动其传播与教育。本研究还获得了知识生成项目的部分资助，该项目由西班牙科学、创新与大学部资助，项目编号为 PID2023-150070NB-I00。

作者感谢 Miguel Cabrera 和 Juan F. Mena 进行了有益的讨论并提出了深刻的见解。

参考文献

Aronszajn, N. (1950). 再生核理论. 美国数学学会会刊, 68(3), 337–404. doi: 10.2307/1990404 Berlyand, L., & Jabin, P. E. (2023). 深度学习数学：导论. Walter de Gruyter. doi: 10.1515/9783111025551 Bohn, B., Griebel, M., & Rieger, C. (2018). 深度核学习的表示定理. arXiv:1709.10441. Brown, L., & Ito, T. (1976). 某些巴拿赫空间预对偶的唯一性. 以色列数学杂志, 24(3), 234–243. doi: 10.1007/BF02834754 Conway, J. B. (1994). 泛函分析教程. Springer. Dales, H. G., & Ülger, A. (2025). 巴拿赫函数代数、Arens正则性与bse范数. Springer. Engel, K.-J., & Nagel, R. (2000). 线性演化方程的单参数半群 (Vol. 194). Springer. doi: 10.1007/b97696 Follain, B., & Bach, F. (2025). 通过正则化增强特征学习：融合神经网络与核方法. 机器学习研究杂志, 26(172), 1–56. Hadamard, J. (1923). 线性偏微分方程中柯西问题讲义. 耶鲁大学出版社. Heeringa, T. J., Spek, L., & Brune, C. (2025). 深度网络是再生核链. arXiv:2501.03697. Jacot, A., Gabriel, F., & Hongler, C. (2018). 神经切线核：神经网络中的收敛与泛化. 载于《神经信息处理系统进展》 (Vol. 31). Kadets, V., & Iacob, A. (2018). 泛函分析与测度论教程. Springer. Kimeldorf, G., & Wahba, G. (1971). 关于切比雪夫样条函数的一些结果. 数学分析与应用杂志, 33(1), 82–95. doi: 10.1016/0022-247X(71)90184-3 Lin, R., Song, G., & Zhang, H. (2021). 配备 ℓ^1 范数的向量值再生核巴拿赫空间中的多任务学习. 复杂性杂志, 63, 101514. doi: 10.1016/j.jco.2020.101514 Lin, R., Zhang, H., & Zhang, J. (2022). 关于再生核巴拿赫空间：通用定义与统一构造框架. 数学学报（英文版）, 38. doi: 10.1007/s10114-022-1397-7 Megginson, R. E. (1998). 巴拿赫空间理论导论. Springer. Naimark, M. A. (1972). 赋范代数. Wolters–Noordhoff. Neyshabur, B., Bhojanapalli, S., & Srebro, N. (2018). 用于神经网络谱归一化间隔界的PAC-贝叶斯方法. 载于《国际学习表征会议论文集》. Pardalos, P. M., Migdalas, A., & Pitsoulis, L. (2008). 帕累托最优性、博弈论与均衡. Springer. Paulsen, V. I., & Raghupathi, M. (2016). 再生核希尔伯特空间理论导论. 剑桥大学出版社. Rao, T. S. S. R. K. (2009). 关于其对偶与μ等距的算子空间(L^1). 算子与矩阵, 3, 397–400. Rudin, W. (1991). 泛函分析. 麦格劳-希尔. Schölkopf, B., Herbrich, R., & Smola, A. J. (2001). 广义表示定理. 载于《国际计算学习理论会议》 (pp. 416–426). Schölkopf, B., & Smola, A. J. (2002). 核学习：支持向量机、正则化、优化及其他. 麻省理工学院出版社. Shenouda, J., Parhi, R., Lee, K., & Nowak, R. D. (2024). 多输出神经网络的变差空间. 机器学习研究杂志, 25(231), 1–40. Sun, H., Gatmiry, K., Ahn, K., & Azizan, N. (2023). 通过镜像下降控制隐式正则化的统一方法. 机器学习研究杂志, 24(393), 1–58. van Neerven, J. (2022). 泛函分析. 剑桥大学出版社. doi: 10.1017/9781009232487 Wang, R., & Xu, Y. (2021). 巴拿赫空间中的表示定理. 机器学习研究杂志, 22(225), 1–65. Wang, R., Xu, Y., & Yan, M. (2024). 用于再生核巴拿赫空间学习的稀疏表示定理. 机器学习研究杂志, 25(93), 1–45. Wang, R., Xu, Y., & Yan, M. (2025). 深度学习的假设空间. arXiv:2403.03353. Wojtowytsch, S. (2020). 与多层ReLU网络相关的巴拿赫空间：函数表示、逼近理论与梯度下降动力学. 中国工业与应用数学学会会刊, 1, 387– 440. doi: 10.4208/csiam-am.20-211 Xu, Y. (2023). 巴拿赫空间中的稀疏机器学习. 应用数值数学, 187, 138–157. doi: 10.1016/j.apnum.2023.02.011 Zhang, H., Xu, Y., & Zhang, J. (2009). 用于机器学习的再生核巴拿赫空间. 机器学习研究杂志, 10, 2741–2775.