arXiv:2304.02220v2 [cs.LG] 8 Apr 2023

关于径向基函数神经网络的通用逼近性质

Aysu Ismayilova
爱丁堡大学数学学院，Peter Guthrie Tait 路，爱丁堡
EH9 3FD，苏格兰，电子邮箱：
a.ismayilova-1@sms.ed.ac.uk

Muhammad Ismayilov法阿大学，尼扎米街
183 号，巴库，阿塞拜疆，电子邮
箱：mv.ismayilov@gmail.com

摘要。本文研究一类新型径向基函数（RBF）神经网络，其中平滑因子被平移所取代。我们证明，在激活函数满足某些条件时，这类网络能够在 d 维欧几里得空间的任意紧子集上逼近任意连续多元函数。对于具有有限个固定中心的径向基函数网络，我们给出了保证实现任意精度逼近的条件。

关键词：径向基函数神经网络、激活函数、均值周期函数、中心、平移

2010年数学学科分类：41A30、41A63、68T05、92B20

1. 引言

径向基函数（Radial Basis Function）神经网络正被用于函数逼近、时间序列预测、分类、模式识别及系统控制问题。除了具备强大的逼近能力，这类网络还具有其他卓越特性，例如能够表示复杂的非线性映射，并以较低的计算成本提供快速而稳健的学习机制。关于径向基函数神经网络的各个方面及其多种应用，已有大量文献论述。

径向基函数神经网络的基本原理和优点最早见于 Broomhead 和 Lowe [2], Moody 和 Darken [9], Lipmann [7] 以及 Bishop[1] 的论文。下面介绍一种具有输入层、隐藏层和输出层的径向基函数网络的构造方案。该网络隐藏层中的每个单元都有自己的中心；对于输入向量 $\mathbf{x} = (x_1, ..., x_d)$，它计算 $\mathbf{x}$ 与其中心 $\mathbf{c} \in \mathbb{R}^d$ 之间的距离。其输出（即给定隐藏单元的输出）是该距离的某个非线性函数。因此，每个隐藏单元都计算一个径向函数，也就是说，一个在球面 $\|\mathbf{x} - \mathbf{c}\| = \alpha$, $\alpha \in \mathbb{R}$ 上为常数的函数。每个输出单元都对隐藏单元的输出进行加权求和。为便于阐述，以下我们仅考虑一维输出空间，而不考虑由多个单元表示的输出

。将我们的结果推广到 n 维输出空间是显而易见的。

假设有 d 个输入单元和一个输出单元，最终响应函数具有以下形式：

$$G(\mathbf{x}) = \sum_{i=1}^{m} w_i g\left(\frac{\|\mathbf{x} - \mathbf{c}_i\|}{\sigma_i}\right). \tag{1.1}$$

这里，$m \in \mathbb{N}$ 是隐藏层中的单元数，$(w_1, ..., w_m) \in \mathbb{R}^m$ 是权重向量，$\mathbf{x} \in \mathbb{R}^d$ 是输入向量，$\mathbf{c}_i \in \mathbb{R}^d$ 和 $\sigma_i \in \mathbb{R}_+$ 分别是第 i个节点 $1 \leq i \leq m$的中心和宽度（或平滑因子），$\|\mathbf{x} - \mathbf{c}_i\|$ 表示$\mathbf{x}$与$\mathbf{c}_i$之间的欧几里得距离，而 $g: [0, +\infty) \to \mathbb{R}$ 是所谓的激活函数。

径向基函数神经网络可以采用各种激活函数，平滑因子可以相同，也可以因单元而异。

径向基函数神经网络具有通用逼近性质。从理论上讲，如果恰当选择激活函数，这类网络可以以任意精度逼近任何连续多元函数。最著名的结果出自 Park 和 Sandberg [10]。1993年，他们证明了，对于连续且可积的 $g(\|\mathbf{x}\|)$（视为 d 变量的函数），函数（1.1）的集合在 $\mathbb{R}^d$上的紧子集上一致收敛拓扑下，在 $C(\mathbb{R}^d)$中稠密。也就是说，对于任意连续函数 $f: \mathbb{R}^d \to \mathbb{R}$、任意紧子集$K \subset \mathbb{R}^d$ 以及任意 $\varepsilon > 0$，都存在一个具有形式（1.1）的函数 G，使得

$$\|f - G\|_K \stackrel{def}{=} \max_{\mathbf{x} \in K} |f(\mathbf{x}) - G(\mathbf{x})| < \varepsilon.$$

Liao、Fang 和 Nuttle [6] 放宽了 $g(\|\mathbf{x}\|)$ 的可积性要求。他们证明，对于几乎处处连续、局部本质有界且非多项式的激活函数，径向基函数网络 (1.1) 可以以任意精度逼近任意连续函数。关于径向基函数神经网络的普适性还有其他结果（例如，参见 [4, 6, 11, 17]）。

本文研究一类新的径向基函数神经网络。在这一类网络中，平滑因子 σ_i 被平移 $\nu_i \in \mathbb{R}$ 取代。也就是说，这一类网络由形如 $\mathbb{R}^d \to \mathbb{R}$ 的函数 H 构成。

$$H(\mathbf{x}) = \sum_{i=1}^{m} w_i g\left(\|\mathbf{x} - \mathbf{c}_i\| - \nu_i\right). \tag{1.2}$$

我们关注此类径向基函数神经网络的通用逼近性质。对于哪些激活函数 g，形式为 (1.2) 的函数在 $\mathbb{R}^d$ 的任意紧子集上按一致收敛拓扑在 $C(\mathbb{R}^d)$ 中稠密？我们将给出激活函数 g 的若干条件，以保证对于任意紧集 $X \subset \mathbb{R}^d$，形式为 (1.2) 的函数在 $C(X)$ 中稠密。

2. 通用逼近定理

以下　　Wing 定理基于 Park 和 Sandberg 以及 Schwartz 的研究结果。　　.

定理 2.1。假设 $d \geq 1$, $1 \leq p < \infty$. 设激活函数 $g \in C(\mathbb{R}) \cap L^p(\mathbb{R})$ 且 $t^{d-1}g(t)$ 在 $[0, +\infty)$ 上可积。那么，对于任意连续函数$f: \mathbb{R}^d \to \mathbb{R}$、任意 $\varepsilon > 0$ 以及任意紧子集 $X \subset \mathbb{R}^d$，存在 $m \in \mathbb{N}$,$w_i, \nu_i \in \mathbb{R}$, $\mathbf{c}_i \in \mathbb{R}^d$，使得

$$\left| f(\mathbf{x}) - \sum_{i=1}^{m} w_i g\left(\|\mathbf{x} - \mathbf{c}_i\| - \nu_i\right) \right| < \varepsilon \tag{2.1}$$

对所有 $\mathbf{x} \in X$.

证明。Park 和 Sandberg 的结果（见 [10, 定理 5]）指出，若 $K: \mathbb{R}^d \to \mathbb{R}$连续、可积，且关于欧几里得范数径向对称，则下列形式的函数

$$q(\mathbf{x}) = \sum_{i=1}^{k} w_i K\left(\frac{\mathbf{x} - \mathbf{c}_i}{\sigma_i}\right). \tag{2.2}$$

对任意紧集 $X \subset \mathbb{R}^d$，在 $C(X)$ 中稠密。

考虑函数 $K(\mathbf{x}) = g(\|\mathbf{x}\|)$。显然，$K(\mathbf{x})$ 是径向对称的，并且由于 $t^{d-1}g(t)$ 在 $[0, +\infty)$ 上可积，$K(\mathbf{x})$ 在 $\mathbb{R}^d$ 上可积。因此，该函数满足 Park 和 Sandberg 定理的假设。假设给定任意函数 $f \in C(\mathbb{R}^d)$、任意数值 $\varepsilon > 0$ 以及任意紧子集 $X \subset \mathbb{R}^d$。根据 Park 和 Sandberg 的上述结果，存在 $k \in \mathbb{N}$、$w_i \in \mathbb{R}$、$\sigma_i > 0$、$\mathbf{c}_i \in \mathbb{R}^d$，使得

$$\left| f(\mathbf{x}) - \sum_{i=1}^{k} w_i g\left(\frac{\|\mathbf{x} - \mathbf{c}_i\|}{\sigma_i}\right) \right| < \frac{\varepsilon}{2} \tag{2.3}$$

对所有 $\mathbf{x} \in X$ 均成立。注意，我们可以将不等式 (2.3) 写成如下形式

$$\left| f(\mathbf{x}) - \sum_{i=1}^{k} g_i\left(\|\mathbf{x} - \mathbf{c}_i\|\right) \right| < \frac{\varepsilon}{2}, \tag{2.4}$$

其中 $g_i(t) := w_i g(t/\sigma_i)$ $t \in \mathbb{R}$。由于 X 是紧的，且距离函数 $\|\cdot\|$ 连续，集合 $\{\|\mathbf{x} - \mathbf{c}_i\| : \mathbf{x} \in X\}$ 是 $\mathbb{R}$ 的紧子集；因此，对于某个有限的 a_i 和 b_i、$i = 1$、...、k，有 $\{\|\mathbf{x} - \mathbf{c}_i\| : \mathbf{x} \in X\} \subset [a_i, b_i]$。

1947 年，Schwartz [13] 证明，连续且属于 p 次（$1 \leq p < \infty$）Lebesgue 可积的单变量函数不是均周期函数（另见 [12, 命题 3.12]）。注意，若函数 $u \in C(\mathbb{R}^d)$ 满足集合 $span\{u(\mathbf{x} - \mathbf{a}) : \mathbf{a} \in \mathbb{R}^d\}$ 在紧集上一致收敛拓扑下不在 $C(\mathbb{R}^d)$ 中稠密（见 [13]），则称其为均周期函数。由于$g \in C(\mathbb{R}) \cap L^p(\mathbb{R})$，根据 Schwartz 的这一结果，集合

$$span\ \{g(t-\lambda):\ \lambda \in \mathbb{R}\}$$

在一致收敛拓扑下于 $C(\mathbb{R})$ 中稠密。这个稠密性结果意味着，对于给定的 ε，存在数 ρ_{ij}、$\lambda_{ij} \in \mathbb{R}$、$i = 1$、2、.....、$k$、$j = 1$、.....、$s_i$，使得

$$\left| g_i(t) - \sum_{j=1}^{s_i} \rho_{ij} g(t - \lambda_{ij}) \right| < \frac{\varepsilon}{2k} \tag{2.5}$$

对所有 $t \in [a_i, b_i]_i,\ i = 1, 2, ..., k.$ 由 (2.4) 和 (2.5) 可得

$$\left| f(\mathbf{x}) - \sum_{i=1}^{k} \sum_{j=1}^{s_i} \rho_{ij} g(\|\mathbf{x} - \mathbf{c}_i\| - \lambda_{ij}) \right| < \varepsilon, \tag{2.6}$$

对所有 $\mathbf{x} \in X.$ 将 (2.6) 中的双重求和改写为单重求和后，即可得到 (2.1) 成立。

推论 2.1. 假设 g 是定义在 $\mathbb{R}$ 上的连续、单调且有界函数，并且 $t^{d-1}g(t)$ 在 $[0, +\infty$ 上可积。那么，对于任意连续函数 $f: \mathbb{R}^d \to \mathbb{R}$、任意 $\varepsilon > 0$ 以及任意紧子集 $X \subset \mathbb{R}^d$，存在 $m \in \mathbb{N}, w_i, \nu_i \in \mathbb{R}, \mathbf{c}_i \in \mathbb{R}^d$，使得

$$\left| f(\mathbf{x}) - \sum_{i=1}^{m} w_i g\left(\|\mathbf{x} - \mathbf{c}_i\| - \nu_i\right) \right| < \varepsilon$$

对所有 $\mathbf{x} \in X.$

证明. 在 [3] 中，Funahashi 证明了：若 g 是定义在 $\mathbb{R}$ 上的连续、单调且有界函数，则对于任意实数 α，函数 $h(t) = g(t + \alpha) - g(t - \alpha)$ 属于 $L_1(\mathbb{R})$。因此，函数 $h(t)$ 不是均周期函数，因为它是连续且 L_1 的函数。此外，$t^{d-1}h(t)$ 在 $[0, +\infty$ 上可积。我们可以将上述定理应用于 h，然后将 $h(\cdot)$ 改为 $g(\cdot + \alpha) - g(\cdot - \alpha)$，即可得到所需结果。

在下述定理中，不要求满足可积性条件。

定理 2.2. 假设 g 是定义在 $\mathbb{R}$ 上的非常值连续有界函数，并且其在无穷远处或负无穷远处存在极限。那么，对于任意连续函数 $f: \mathbb{R}^d \to \mathbb{R}$、任意 $\varepsilon > 0$ 以及任意紧子集 $X \subset \mathbb{R}^d$，存在 $m \in \mathbb{N}, w_i, \nu_i \in \mathbb{R}, \mathbf{c}_i \in \mathbb{R}^d$，使得

$$\left| f(\mathbf{x}) - \sum_{i=1}^{m} w_i g\left(\|\mathbf{x} - \mathbf{c}_i\| - \nu_i\right) \right| < \varepsilon$$

对所有 $\mathbf{x} \in X.$

证明。 g 上的条件意味着函数 $K_0(\mathbf{x}) = g(\|\mathbf{x}\|)$ 是一个非常数、连续且有界的多元函数。因此， $K_0(\mathbf{x})$ 不是多项式。这里我们使用 Liao、Fang 和 Nuttle 在 [6], 中的结果。他们证明，对于一个几乎处处连续、局部本质有界且非多项式的函数 $K(\mathbf{x})$，径向基函数网络 $\sum_{i=1}^{m} w_i K\left(\frac{\mathbf{x}-\ i}{\sigma_i}\right)$ c在任意紧集 $X \subset \mathbb{R}^d$ 上的 $C(X)$ 中稠密。注意，我们的函数 $K_0(\mathbf{x})$ 满足这一结果的条件。

此外，由 Schwartz 的一个定理（参见 [13, 第907] 页和 [12, 命题 3.12]）可知， g 不是均周期的。因此，对于任意闭区间 $[a, b].$，集合 $span\{g(t-\lambda): \ \lambda \in \mathbb{R}\}$ 在 $C[a, b]$ 中稠密。现在，我们已经掌握了所有必要的事实，可以重复上述定理 2.1 证明中的思路，从而得到所需的最终结果。

3. 具有有限个质心的径向基函数网络

本节研究具有有限个固定中心的径向基函数神经网络的逼近性质。我们描述一些紧集 $X \subset \mathbb{R}^d$，在这些集合上，此类网络在 $C(X)$ 中稠密。

假设给定 k 个固定中心 $\mathbf{c}_1$， ...， $\mathbf{c}_k$。令 $S = \{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$。考虑仅具有这些中心且平移任意的径向基函数网络集合

$$\mathcal{G}(g, S) = \left\{ \sum_{i=1}^{m} w_i g\left(\|\mathbf{x} - \mathbf{c}\| - \nu_i\right) : \mathbf{c} \in S, \ w_i, \nu_i \in \mathbb{R}, \ m \in \mathbb{N} \right\}.$$

在上述集合中，我们只固定集合 $S = \{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$ 和激活函数 $g \in C(\mathbb{R})$。自然会产生以下问题：

1) $\mathcal{G}(g, S)$ 中的径向基函数网络在紧集上一致收敛拓扑下是否在 $C(\mathbb{R}^d)$ 中稠密？也就是说，对于任意紧集 X， $\mathcal{G}(g, S)$ 是否 $= C(X)$?

2) 如果上述问题的答案是否定的，那么对于哪些紧集 $X \subset \mathbb{R}^d$， $\mathcal{G}(g, S) = C(X)$?

请注意，第一个问题的答案确实是否定的。为此，引入以下函数集：

$$\mathcal{R}(S) = \left\{ \sum_{i=1}^{k} g_i\left(\|\mathbf{x} - \mathbf{c}_i\|\right) : g_i \in C(\mathbb{R}) \right\}.$$

在这个集合中， $S = \{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$ 是固定的，我们改变连续函数 $g_i.$。注意，这是一个线性空间。由于 $\mathcal{G}(g, S)$ 中的每一项 $w_i g(\|\mathbf{x} - \mathbf{c}\| - \nu_i)$ 都是某个 $\mathbf{c}_i$ 对应的 $g_i(\|\mathbf{x} - \mathbf{c}_i\|)$ 形式的函数，因此我们可以推得 $\mathcal{G}(g, S) \subset \mathcal{R}(S)$。因此，具有固定质心的径向基函数网络集合小于具有这些质心的径向函数线性组合集合。因此，如果 $\mathcal{G}(g, S)$ 在 $C(X)$ 中稠密，那么集合 $\mathcal{R}(S)$ 也将稠密。但遗憾的是，对于极多紧集 X， $\mathcal{R}(S)$ 在 $C(X)$ 中并不稠密。这里缺乏稠密性的原因与下面的定理有关，该定理

归功于维图什金和亨金 [16]: 对于定义在立方体 $[a,b]^d$ 上的任意 k 个固定连续可微函数 h_i、$i=1$、...、k，函数集合

$$\left\{\sum_{i=1}^{k} g_i\left(h_i(x_1,...,x_d)\right) : g_i \in C(\mathbb{R})\right\}$$

在一致收敛拓扑下，它在 $[a,b]^d$ 上所有连续函数组成的空间中无处稠密。因此，如果涉及所有紧集 X，那么 $\mathcal{G}(g,S)$ 就不能在 $C(X)$ 中稠密。例如，由于任何具有内点的集合都包含一个足够小的立方体 $[a,b]^d$，由维图什金和亨金的结果可知，对于任何具有内点的紧集 X，$\mathcal{G}(g,S)$ 在 $C(X)$ 中都不稠密。不过，仍可能存在某些紧集 X，使得 $\mathcal{G}(g,S)=C(X)$（例如，~~取单点集~~ $X=\{\mathbf{x}\}$）。如何刻画这类集合？为回答这个问题，我们引入以下称为循环的对象：

定义 3.1. 点集 $l=\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}\subset\mathbb{R}^d$ 称为相对于质心 $\mathbf{c}_1,...,\mathbf{c}_k$ 的循环，如果存在一个向量 $\lambda=(\lambda_1,\ldots,\lambda_n)\in\mathbb{Z}^n\setminus\{\mathbf{0}\}$，使得

$$\sum_{j=1}^{n}\lambda_j\delta_{\|\mathbf{x}_j-\mathbf{c}_i\|}(t)=0,\quad for\ all\ i=1,\ldots,k. \tag{3.1}$$

在上述定义中，$\delta_{\|\mathbf{x}_j-\mathbf{c}_i\|}(t)$ 是单点集$\{\|\mathbf{x}_j-\mathbf{c}_i\|\}$的示性函数。也就是说，

$$\delta_{\|\mathbf{x}_j-\mathbf{c}_i\|}(t)=\begin{cases}1, & \text{if } t=\|\mathbf{x}_j-\mathbf{c}_i\|\\ 0, & \text{if } t\neq\|\mathbf{x}_j-\mathbf{c}_i\|\end{cases}.$$

让我们更仔细地考察公式（3.1）。我们将看到，它实际上表示一个简单的线性方程组。为理解这一点，固定下标 $i.$。设集合 $\{\|\mathbf{x}_j-\mathbf{c}_i\|,j=1,...,n\}$ 中有 s_i 个不同的值，我们将它们记为 $\gamma_1^i,\gamma_2^i,...,\gamma_{s_i}^i$。取第一个数$\gamma_1^i.$，令 $t=\gamma_1^i$，由（3.1）可得

$$\sum_j \lambda_j=0,$$

其中，求和遍历所有满足 j 的 $\|\mathbf{x}_j-\mathbf{c}_i\|=\gamma_1^i.$。这是关于 $\lambda_1,...,\lambda_n.$ 的第一个线性方程。该方程对应于 γ_1^i。现在取 γ_2^i。同样地，将$t=\gamma_2^i$ 代入（3.1），即可得到第二个方程。依此继续，直到 $\gamma_{s_i}^i$，我们得到关于 $\lambda_1,...,\lambda_n$的 s_i 个齐次线性方程。这些方程的系数是整数 0 和 1。改变 i，最终得到 $s=\sum_{i=1}^{k}s_i$ 个这样的方程。因此可见，（3.1）的展开形式表示由这些线性方程组成的方程组。因此，如果该方程组有一个非零整数分量解，则集合$l=\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ 构成一个循环。事实上，不难理解，如果方程组（3.1）有一个非零实数分量解，那么它也有一个非零整数分量解。这意味着在上述定义中，我们可以用 $\mathbb{R}^n\setminus\{\mathbf{0}\}$替换 $\mathbb{Z}^n\setminus\{\mathbf{0}\}$。

这里给出两个简单的循环示例。读者很容易再举出许多其他例子。设在 xy 平面上给定两个质心 $\mathbf{c}_1 = (0, 0)$ 和 $\mathbf{c}_2 = (4, 0)$。那么，直线 $x = 2$上的任意两点 A 和 B，只要它们关于直线 $y = 0$对称，就构成一个循环。事实上， A 和 B 到 $\mathbf{c}_1$ 的距离相等；在 $i = 1$ 的情况下，公式（3.1）为 $\lambda_1 + \lambda_2 = 0.$。由于 A 和 B 到 $\mathbf{c}_2$ 的距离也相等，公式（3.1）对 $i = 2.$ 给出相同的方程。因此， $\{A, B\}$ 是一个2点循环，向量（λ_1, λ_2）可取为（$-1, 1$）。同样也容易构造一个相对于这些质心的4点循环。按给定顺序，考虑 xy 平面上的四个圆： $\|\mathbf{x} - \mathbf{c}_1\| = 2$、 $\|\mathbf{x} - \mathbf{c}_1\| = 3$、 $\|\mathbf{x} - \mathbf{c}_2\| = 4$、$\|\mathbf{x} - \mathbf{c}_2\| = 3$。这些圆在 xy 平面的第一象限内交于4个点 A, B, C, D。每个圆只经过这四个点中的两个，不难验证公式（3.1）可化为如下方程组

$$\begin{cases} \lambda_1 + \lambda_2 = 0 \\ \lambda_3 + \lambda_4 = 0 \\ \lambda_2 + \lambda_3 = 0 \\ \lambda_1 + \lambda_4 = 0 \end{cases}$$

其有一个解（$-1, 1 - 1, 1$）。因此，相对于质心 $\mathbf{c}_1$ 和 $\mathbf{c}_2$， $\{A, B, C, D\}$ 是一个环。

上面的第二个例子启发我们考虑相对于任意给定的两个质心 $\mathbf{c}_1$ 和 $\mathbf{c}_2$ 的一般环。在这一特殊情况下，我们将使用“闭路径”一词，而不是“环”。

定义3.2。假设 $l = (\mathbf{x}_1, \mathbf{x}2, ..., \mathbf{x}_n)$（其中 $\mathbf{x}_i \neq \mathbf{x}_{i+1}$）是一个有序集，并具有以下性质： $\|\mathbf{x}_1 - \mathbf{c}_1\| = \|\mathbf{x}_2 - \mathbf{c}_1\|$、 $\|\mathbf{x}_2 - \mathbf{c}_2\| = \|\mathbf{x}_3 - \mathbf{c}_2\|$、 $\|\mathbf{x}_3 - \mathbf{c}_1\| = \|\mathbf{x}_4 - \mathbf{c}_1\|$， ……，或 $\|\mathbf{x}_1 - \mathbf{c}_2\| = \|\mathbf{x}_2 - \mathbf{c}_2\|$、 $\|\mathbf{x}_2 - \mathbf{c}_1\| = \|\mathbf{x}_3 - \mathbf{c}_1\|$、 $\|\mathbf{x}_3 - \mathbf{c}_2\| = \|\mathbf{x}_4 - \mathbf{c}_2\|$， ……。那么，相对于质心 $\mathbf{c}_1$ 和 $\mathbf{c}_2$， l 被称为一条路径。具有偶数个点的路径（$\mathbf{x}_1, \mathbf{x}2, ..., \mathbf{x}_{2n}$）如果（$\mathbf{x}_1, \mathbf{x}2, ..., \mathbf{x}_{2n}, \mathbf{x}_1$）也是一条路径，则称其为闭路径。

注意，闭路径就是环。事实上，如果 $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{2n})$ 是一条闭路径，那么不难看出，对于一个分量为 $\lambda_i = (-1)^i$ 的向量 $\lambda = (\lambda_1, \ldots, \lambda_{2n})$，我们有

$$\begin{aligned} \sum_{j=1}^{2n} \lambda_j \delta_{\|\mathbf{x}_j - \mathbf{c}_1\|} &= 0, \\ \sum_{j=1}^{2n} \lambda_j \delta_{\|\mathbf{x}_j - \mathbf{c}_2\|} &= 0. \end{aligned}$$

因此，根据定义 3.1，集合 $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{2n}\}$ 相对于中心 $\mathbf{c}_1$ 和 $\mathbf{c}_2$ 构成一个环。

循环和路径不仅可以针对距离函数 $d(\mathbf{x}) = \|\mathbf{x} - \mathbf{c}\|$ 定义，也可以针对其他有用的函数定义。针对内积

a · x 定义的这些对象历史悠久，且已被证明在脊函数理论中非常高效（即形如 g(a · x) 的函数）。例如，可参阅 Ismailov 的专著 [5]。

令 c1, ..., c$_k$ 为固定中心， X 为 $\mathbb{R}^d$的一个紧子集。对于每个 $i = 1, \ldots, k$，考虑以下集合函数

$$\tau_i : 2^X \to X,\ \tau_i(Z) = \{\mathbf{x} \in Z :\ |d_i^{-1}(d_i(\mathbf{x})) \bigcap Z| \geq 2\},$$

其中 d_i(x) $= \|$x $-$ c$_i\|$，符号 $|\ \ |$表示所考虑集合的基数。定义 $\tau(Z)$ 为 $\bigcap_{i=1}^{k} \tau_i(Z)$，并定义 $\tau^2(Z) = \tau(\tau(Z))$、 $\tau^3(Z) = \tau(\tau^2(Z))$，依此类推。显然， $\tau(Z) \supseteq \tau^2(Z) \supseteq \tau^3(Z) \supseteq ...$。有可能对于某个 n， $\tau^n(Z) = \emptyset$。在这种情况下，可以看出 Z 不包含环。一般而言，如果某个集合 $Z \subset X$ 构成一个环，那么 $\tau^n(Z) = Z.$。值得指出的是，集合函数 τ_i 最早出现在 Sternfeld [15], 中；在该处，涉及的是一般连续函数，而非距离函数 d_i(x)。

以下定理成立。

定理 3.1. 设 X 是 $\mathbb{R}^d$ 的紧子集。若 $\cap_{n=1,\ 2,\ ...}\tau^n(X) = \emptyset$，则集合 $\mathcal{R}(S)$ 在 $C(X)$ 中稠密。

由于形如 $g_i(\|$x $-$ c$_i\|)$ 的函数生成空间 $C(X)$ 的一个子代数，定理 3.1 立即由 Sproston 和 Straus [14] 关于连续函数代数之和所证明的一般结果推出。

以下定理给出了径向基函数神经网络 $\mathcal{G}(g, S)$ 在 $C(X)$ 中稠密的充分条件和必要条件。

定理 3.2. 假设 g 是一个连续的 p 次（$1 \leq p < \infty$）可积函数，或者 g 是一个非恒定的连续有界函数，且其在无穷远处（或负无穷远处）有极限。则以下陈述成立：

(a) 若 $\bigcap_{n=1,\ 2,\ ...}\tau^n(X) = \emptyset$，则集合 $\mathcal{G}(g, S)$ 在 $C(X)$ 中稠密；(b) 若 $\mathcal{G}(g, S)$ 在 $C(X)$ 中稠密，则集合 X 不含环（相对于 S）。

证明。(a) 假设 $\bigcap_{n=1,\ 2,\ ...}\tau^n(X) = \emptyset$。根据定理 3.1，集合 $\mathcal{R}(S)$ 在$C(X)$中稠密。因此，对于任意函数 $f \in C(X)$ 和任意正实数 ε ，存在连续函数 g_i、 $i = 1$、...、 k，使得

$$\left| f(\mathbf{x}) - \sum_{i=1}^{k} g_i\left(\|\mathbf{x} - \mathbf{c}_i\|\right) \right| < \frac{\varepsilon}{k+1} \tag{3.2}$$

对所有 x $\in X$。由于 X是紧的，集合 $Y_i = \{\|$x $-$ c$_i\| :$ x $\in X\}$, $i = 1, 2, ..., k$ 也都是紧集。我们从第 2 节得知，函数 g 不是均周期的，因此集合

$$span\ \{g(t - \theta) :\ \theta \in \mathbb{R}\}$$

在紧集上一致收敛拓扑下于 $C(\mathbb{R})$ 中稠密（参见定理 2.1 和 2.2 的证明）。由此可知，对于上述 ε，存在 $c_{ij}, \theta_{ij} \in \mathbb{R}, i = 1, 2, ..., k, j = 1, ..., m_i$，使得

$$\left| g_i(t) - \sum_{j=1}^{m_i} c_{ij} g(t - \theta_{ij}) \right| < \frac{\varepsilon}{k+1} \tag{3.3}$$

对所有 $t \in Y_i,\ i = 1, 2, ..., k.$ 由 (3.2) 和 (3.3) 可得

$$\left| f(\mathbf{x}) - \sum_{i=1}^{k} \sum_{j=1}^{m_i} c_{ij} g(\|\mathbf{x} - \mathbf{c}_i\| - \theta_{ij}) \right| < \varepsilon. \tag{3.4}$$

对所有 $\mathbf{x} \in X$。因此 $\overline{\mathcal{G}(g, S)} = C(X)$。

(b) 假设 $\mathcal{G}(g, S)$ 在 $C(X)$ 中稠密。那么，对于任意正实数 ε，不等式 (3.4) 对某些系数 $c_{ij}, \theta_{ij},\ i = 1, 2, ..., k,\ j = 1, ..., m_i.$ 成立。由于对每个 $i = 1, 2, ..., k$，函数 $\sum_{j=1}^{m_i} c_{ij} g(\|\mathbf{x} - \mathbf{c}_i\| - \theta_{ij})$ 都是形如 $g_i(\|\mathbf{x} - \mathbf{c}_i\|)$ 的函数，由 (3.4) 可知，子空间 $\mathcal{R}(S)$ 在 $C(X)$ 中稠密。下面证明 X 不包含环。反设 X 包含一个环，记为 $l = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$。设 $\lambda = (\lambda_1, \ldots, \lambda_n)$ 为定义 3.1 中给出的向量。引入泛函

$$F : C(X) \to \mathbb{R},\ F(f) = \sum_{j=1}^{n} \lambda_j f(\mathbf{x}_j).$$

显然，F是一个有界线性泛函，其范数为 $\sum_{j=1}^{n} |\lambda_j|$。不难验证，对于任意 $h \in \mathcal{R}(S)$，都有 $F(h) = 0$。根据乌雷松引理，存在连续函数 $f_0 : X \to \mathbb{R}$，使得当 $\lambda_j > 0$时，$f_0(\mathbf{x}j) = 1$；当 $\lambda_j < 0$ 时，$f_0(\mathbf{x}j) = -1$；并且对于任意$\mathbf{x} \in X \backslash l$，$-1 < f_0(\mathbf{x}) < 1$。对于此函数，$F(f_0) \neq 0$。我们构造出了一个非零湮灭泛函 F。此类泛函的存在意味着 $\mathcal{R}(S)$在 $C(X)$中不可能稠密。由此得到的矛盾证明了定理的第二个结论。

最后，我们想指出，对于只有两个固定中心的径向基函数神经网络，其稠密性问题的解具有明确的几何表述。在这一特殊情形下，我们可以完整刻画所有满足以下条件的紧集 $X \subset \mathbb{R}^d$：$\mathcal{G}(g, S)$在$C(X)$中稠密。为表述该定理，考虑 X中点之间的以下关系。若$\mathbf{x}$和$\mathbf{y}$属于 X 中的某条路径，则$\mathbf{x} \sim \mathbf{y}$；这一关系定义了一种等价关系。其等价类称为轨道（见 [8]）。

以下定理成立。

定理 3.3. 假设 g 是一个连续的 p 次（$1 \leq p < \infty$）可积函数，或者 g 是一个非恒定的连续有界函数，且在无穷远处（或负无穷远处）有极限。假设 X 是 $\mathbb{R}^d$ 的一个紧子集，其中所有轨道都是闭的，并且 $S = \{\mathbf{c}_1, \mathbf{c}_2\}$ 是固定中心的集合。那么，当且仅当 X 不包含闭路径时，集合 $\mathcal{G}(g, S)$ 在 $C(X)$ 中稠密。

证明可以用与前一个定理相似的方法进行。只需将定理 3.1 替换为以下结果；该结果是 Marshall 和 O’Farrell [8] 关于两个函数代数之和的一致逼近的一般结果的推论：如果 X 的所有轨道都是闭的，那么，要使 $\mathcal{R}(S)$ 在 $C(X)$ 中稠密，必要且充分的条件是 X 不包含闭路径。

备注。根据定义，闭路径是某个点在球面 $\|\mathbf{x}-\mathbf{c}_1\|=r_1,\ \|\mathbf{x}-\mathbf{c}_2\|=r_2$（$r_1$ 和r_2 不固定）上交替地从一个位置跳到另一个位置，最后返回其初始位置的轨迹。在 $\mathbb{R}^2$ 中，圆 $\|\mathbf{x}-\mathbf{c}_1\|=r_1,\ \|\mathbf{x}-\mathbf{c}_2\|=r_2$ 构成一个圆形网格。因此，要使 G（g, S）在 C（X）中稠密，充分必要条件是 X 不包含该网格中任何顶点（交点）序列 $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \mathbf{x}_1\}$，并且满足这一前提：数对 $\mathbf{x}_i, \mathbf{x}_{i+1}$ 和 $\mathbf{x}_{i+1}, \mathbf{x}_{i+2}$ 位于不同的圆上。

参考文献

[1] C. Bishop，《改进径向基函数神经网络的泛化性能》，Neural Computation，3(4)（1991年），579–588。[2] D.S. Broomhead、D. Lowe，《多变量函数插值与自适应网络》，Complex Systems，2（1988年），321–355。[3] K. Funahashi，《神经网络对连续映射的近似实现》，Neural Networks，2（1989年），183–192。[4] G. B. Huang、L. Chen 和 C. K. Siew，《使用带有随机隐藏节点的增量构造型前馈网络实现通用逼近》，IEEE Transactions on Neural Networks，17（2006年），第4期，879–892。[5] V. E. Ismailov，《岭函数及其在神经网络中的应用》。Mathematical Surveys and Monographs，263。美国数学会，普罗维登斯，罗得岛州，2021年，186页。[6] Y. Liao、S. C. Fang 和 H. L.W. Nuttle，《径向基函数网络成为通用逼近器的宽松条件》，Neural Networks，16（2003年），第7期，1019–1028。[7] R. P. Lippman，《使用神经网络进行模式分类》，IEEE Communications Magazine，27（1989年），47–64。[8] D. E. Marshall 和 A. G. O’Farrell，《实函数的一致逼近》，Fundamenta Mathematicae，104（1979年），203–211。[9] J. Moody、C. Darken，《利用局部感受野进行学习》。载于：1988年 Connectionist Models Summer School 论文集，Morgan-Kaufmann 出版商，1988年。

[10] J. Park 和 I. W. Sanberg，《逼近与径向基函数网络》，Neural Computation，5（1993年），305–316。[11] J. Park 和 I. W. Sanberg，《利用径向基函数网络实现通用逼近》，Neural Computation，2（1991年），246–257。[12] A. Pinkus，《神经网络中多层感知器模型的逼近理论》，Acta Numerica，8（1999年），143–195。[13] L. Schwartz，《周期平均函数的一般理论》，Ann. Math.，48（1947年），857–928。[14] J. P. Sproston 和 D. Strauss，《$C(X)$ 的子代数之和》，J. London Math. Soc.，45（1992年），265–278。[15] Y. Sternfeld，《一致分离的函数族》，Israel J. Math.，29（1978年），61–91。[16] A. G. Vitushkin、G. M. 亨金，《函数的线性叠加》（俄文），Uspehi Mat. Nauk，22（1967年），77–124。[17] Y. Wu、H. Wang、B. Zhang、K. L. Du，《使用径向基函数网络进行函数逼近与分类》，ISRN Appl. Math，2012（2012年），1–34。