LLM技术洞察 02
多元函数的可表示性定理

Jiangsheng Yu

05/08/2024

历史背景： Kolmogorov-Arnold 表示

1955 年，苏联数学大师**安德雷 · 柯尔莫哥洛夫** (Andrey Kolmogorov, 1903—1987) 主持了一次学生研讨会，聚焦于多元函数（multivariable function，也称“多变量函数”）的近似表示理论[149]。在演讲中，柯尔莫哥洛夫还“将希尔伯特第 13 问题表述为一个几乎肯定不会实现的长期计划”†。次年，柯尔莫哥洛夫证明了任何 n 元函数（$n \geqslant 4$）都可以用有限数量的三元函数来构造。1957 年，他的学生、年仅 19 岁的**弗拉基米尔 · 阿诺尔德** (Vladimir Arnold, 1937—2010) 进一步证明了，每个三元连续函数都可以表示为有限多个二元连续函数的叠加，从而给出了希尔伯特第 13 问题（函数版）的最终解决方案。仅仅几个月后，使用完全不同的构造技术，柯尔莫哥洛夫设法证明了，任意多元连续函数 $f(x_1, \cdots, x_n)$ 都可以表示为若干个一元连续函数的如下“叠加”：

$$f(x_1, \cdots, x_n) = \sum_{q=1}^{2n+1} \chi_q \left[\sum_{p=1}^{n} \phi_{pq}(x_p) \right] \tag{8.1}$$

式 (8.1) 常被称为“柯尔莫哥洛夫-阿诺尔德表示”，简称“KA 表示” (KA representation)。在式 (8.1) 中，内部函数 ϕ_{pq} 是固定的，只有外部函数 χ_q 取决于所表示的函数 f。随后，对 KA 表示有一些改进。

更多的可表示性定理

- 1962 年，美籍俄裔数学家**乔治 · 冈特 · 洛伦兹** (George Gunter Lorentz, 1910—2006) 指出，外部函数 χ_q 可以用单一函数 χ 代替。

- 1965 年，美国数学家**大卫 · 斯普雷彻** (David Sprecher) 将所有内部函数简化为单一函数 ψ 的平移和扩展，满足以下性质：存在 $\epsilon > 0$ 和 $\lambda > 0$，使得任何 n 元连续函数都可以表示为

$$f(x_1, \cdots, x_n) = \sum_{q=1}^{2n+1} \chi \left[\sum_{p=1}^{n} \lambda_{pq} \psi(x_p + \epsilon q) \right] \tag{8.2}$$

式 (8.2) 大大简化了多元连续函数的表示：KA 表示 (8.1) 中的 $n(2n+1)$ 个内部函数 ϕ_{pq} 被单一函数 ψ 和一些常数 λ_{pq} 替代，而 $2n+1$ 个外部函数 χ_q 被单一函数 χ 替代。

- 1967 年，美籍俄裔数学家**布马 · 弗里德曼** (Buma Fridman) 在他的第一篇论文（当时，他还在莫斯科国立大学读本科）证明了，KA 表示 (8.1) 中的内部函数 ϕ_{pq} 可以选择那些满足利普希茨条件 (Lipschitz condition)*的函数。1972 年，斯普雷彻将此结果扩展到表示 (8.2)，即可以选择满足利普希茨条件的函数 ψ 来构造式 (8.2)。

更多的可表示性定理

- 1976 年，美国数学家**拉乌夫・多斯** (Raouf Doss) 发展了 KA 表示，他证明了定义在单位超正方体 $[0,1]^n$ 上的任意连续函数 $f(x_1,\cdots,x_n)$ 皆有以下表示：

$$f(x_1,\cdots,x_n)=\sum_{q=1}^{2n+1} g\left(\prod_{p=1}^{n}\phi_{pq}(x_p)\right),\ \text{一元函数 } g \text{ 具有绝对连续的傅立叶级数} \tag{8.3}$$

结果 (8.3) 的一个推论是，定义在单位正方形 $[0,1]^2$ 上的任意连续函数 $h(x,y)$ 皆有以下表示：

$$h(x,y)=\sum_{k=1}^{\infty} a_k \exp\{\mathrm{i} f_k(x) g_k(y)\},\ \text{其中 } f_k, g_k \text{ 都是一元连续函数且 } \sum_{k=1}^{\infty}|a_k|<\infty$$

多元函数的可表示性定理 (representability theorems) 为机器学习提供了理论支持。特别地，KA 表示 (8.1) 和斯普雷彻表示 (8.2) 已成为人工神经网络的数学基础，其中的多元函数 $f(x_1,\cdots,x_n)$ 可以是定义在 $\mathbb{R}^n$ 的任何给定紧集 (compact set)* 上的任意连续函数。

KAN: Kolmogorov–Arnold Networks

这篇论文的主要思想是提出一种新型神经网络结构，即Kolmogorov-Arnold网络（KAN）作为多层感知器（MLP）的一个有前途的替代方案。KANs的关键创新是在边上使用可学习的激活函数，而不是传统的在节点上使用固定激活函数。此外，KANs没有线性权重矩阵，每一个权重参数都被一个参数化为样条的一元函数所替代。这种结构改变带来了几个显著的优势：

1. **准确性提高**：相较于传统MLP，即使KANs的结构更小，也能在数据拟合和解决偏微分方程（PDE）问题上达到相当或更好的精度。
2. **可解释性增强**：由于KANs的结构和功能的可视化，使得它们在与用户的交互中更加直观和易于理解。
3. **科学发现的潜力**：通过数学和物理的实例展示，KANs能够作为科学家的“合作者”，帮助他们发现或重新发现数学和物理定律。

Kolmogorov-Arnold 可表示性定理

- Kolmogorov-Arnold 可表示性定理是函数论中的一个基本结果，并对神经网络领域有重要的影响。该定理表明，任何多变量连续函数都可以表示为只有一个变量的连续函数的叠加。

- 这个定理的意义在于它表明复杂的多维关系可以被简化为更简单的一维形式。这对于神经网络架构的设计具有深远的影响，例如使得网络能够理论上通过一系列更简单的函数来逼近任何连续函数。特别是，它影响了深度学习模型如神经网络的架构和概念理解，这些模型如何能够有效地学习并表示复杂的数据模式。