arXiv:2302.00806v1 [cs.LG] 2 Feb 2023

保持预言机的潜在流

Alexander R oman [1] Roy T. Forestano [*1] Konstantin T. Matchev [*1] Katia Matcheva [*1] Eyup Unlu [*1]

摘要

我们提出了一种深度学习方法，用于在整个带标签数据集中同时发现多个非平凡连续对称性。对称变换及其相应的生成元通过全连接神经网络进行建模，并使用经过特殊构造的损失函数进行训练，以确保所需的对称性。本文的两个新要素是采用降维潜在空间，以及将方法推广到相对于高维预言机具有不变性的变换。我们在 MNIST 手写数字数据集上通过多个示例展示了该方法。

1. 引言

对称性遍布我们周围的世界，存在于各种尺度之上——从标准模型中对亚原子粒子的微观描述，到宇宙的大尺度结构。对称性既是当代科学的指导原则（Gross，1996），也存在于我们的日常互动之中。因此，数据科学的一项基本任务，就是发现、描述并识别给定数据集中存在的对称性。根据诺特定理（Noether，1918），数据中存在连续对称性，意味着存在一条普遍适用的守恒定律；这一定律对于理解系统的行为和演化不可或缺。同时，对称性也能在艺术中带来审美愉悦，并可用于辨认和评价艺术流派及个别艺术家的作品（Barenboim 等人，2021年）。研究对称性及其底层群性质有着悠久的数学传统，而这些性质本身也颇具研究价值（Wigneretal.，1959年）。

许多研究团队从不同角度探索了将机器学习（机器学习）应用于对称性的研究。现有工作主要关注给定的对称性如何反映在数据的学习到的表示中（Iten et al.，2020年；Dillon et al.，2022年），或反映在机器学习架构本身中，例如神经网络（神经网络）的嵌入层（Krippendorf & Syvaeri, 2020年）。

[*]贡献相同 [1]基础理论研究所，物理系，佛罗里达大学，美国佛罗里达州盖恩斯维尔，邮编 32611。通信作者：Konstantin Matchev < matchev@ufl.edu>。

版权所有 © 2023，作者所有。

另一些方案则尝试设计特殊的机器学习架构，例如等变神经网络，从一开始就内置所需的对称性属性（Butter et al., 2018；Kanwar et al., 2020；Bogatskiy et al., 2020；Gong et al., 2022；Bogatskiy et al., 2022；Li et al., 2022；Hao et al., 2022），并测试其性能（Gruver et al., 2022）。将对称性直接纳入机器学习模型后，模型会更经济（就学习到的表示而言）、更易解释，也更易训练。这一研究方向还可以扩展到离散（置换）对称性（Fenton et al., 2022；Shmakov et al., 2022）。

近年来，机器学习技术开始被应用于更为形式化的数学问题，而这类问题传统上属于理论学家的研究领域。例如，深入理解问题中存在的对称性，有助于揭示守恒量（Cranmer 等人，2020年；Liu & Tegmark，2021年），或为更基本的统一图景提供线索（Wu & Tegmark，2019年）。机器学习已被用于发现势的对称性（Krippendorf & Syvaeri，2020年；Barenboim 等人，2021年；Craven 等人，2022年），判断给定的一对输入是否通过某种对称性相关联（Wetzel 等人，2020年），区分尺度不变对称性与共形对称性（Chen 等人，2020年），以及探索弦理论的图景（He，2017年；Carifio 等人，2017年；Ruehle，2020年）。近期研究利用生成对抗网络（GAN）学习保持概率分布不变的变换（Desai 等人，2022年）。机器学习的应用也已延伸至群论这一为对称性提供抽象数学语言的领域。例如，近期研究使用机器学习计算李群不可约表示的张量积和分支规则（Chen 等人，2021年），并获得数据中存在的对称性的李群生成元（Liu & Tegmark，2022年；Moskalev 等人，2022年；Forestano 等人，2023年）。

本文的主要目标是设计一种深度学习方法，模仿传统理论学家的思维方式，能够从第一性原理出发，在给定数据集中发现并分类全部（连续）对称性，即不依赖任何先验假设或偏见。我们的研究

对相关的先前工作进行了补充和拓展（Krip-pendorf & Syvaeri，2020年；Barenboim 等人，2021年；Liu & Tegmark，2022年；Craven 等人，2022年；Moskalev 等人，2022年；Forestano 等人，2023）。我们的方法具有普适性，无需预先了解数据集中可能存在何种对称性——相反，这些对称性是从零开始学习得到的。

2. 问题定义

我们的出发点是一个带标签的数据集，其中包含 m 个样本，每个样本有 n 个特征和 k 个目标：

$$\begin{array}{cccccccc} x_1^{(1)}, & x_1^{(2)}, & \ldots, & x_1^{(n)}; & y_1^{(1)}, & y_1^{(2)}, & \ldots, & y_1^{(k)} \\ x_2^{(1)}, & x_2^{(2)}, & \ldots, & x_2^{(n)}; & y_2^{(1)}, & y_2^{(2)}, & \ldots, & y_2^{(k)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_m^{(1)}, & x_m^{(2)}, & \ldots, & x_m^{(n)}; & y_m^{(1)}, & y_m^{(2)}, & \ldots, & y_m^{(k)} \end{array} \tag{1}$$

我们使用粗体向量记法

$$\mathbf{x} \equiv \{x^{(1)}, x^{(2)}, \ldots, x^{(n)}\} \in \mathbb{R}^n \tag{2}$$

表示 n 维输入特征，并使用箭头向量记法表示 k 维目标向量

$$\vec{y} \equiv \{y^{(1)}, y^{(2)}, \ldots, y^{(k)}\} \in \mathbb{R}^k. \tag{3}$$

该数据集 (1) 可简洁地写为$\{\mathbf{x}_i; \vec{y}_i\}$，其中 $i = 1,\ 2,\ \ldots\ldots,\ m$。

为了定义数据的对称性（1），我们将利用向量函数 $\vec{\varphi}(\mathbf{x})$，它充当预言机，生成相应的目标标签 $\vec{y}_1,\ \vec{y}_2,\ \ldots\ldots,\ \vec{y}_m$：

$$\vec{y}_i \ \equiv \ \vec{\varphi}(\mathbf{x}_i)\,, \qquad i \ = \ 1, 2, \ldots, m\,. \tag{4}$$

这是给定数据集上的理想分类器。在数据科学应用中（这是本文主要关注的领域），需要通过标准回归方法从数据集（1）中数值地学习向量预言机 $\vec{\varphi}$：$\mathbb{R}^n \to \mathbb{R}^k$。另一方面，在某些较为形式化的理论应用中，预言机 $\vec{\varphi}$ 也可以由外部直接以解析函数的形式提供。本文提出的方法适用于这两种情况。

在此设定下，主要目标是推导出一个对称变换$\mathbf{f}$：$\mathbb{R}^n \to \mathbb{R}^n$

$$\mathbf{x}' \ = \ \mathbf{f}(\mathbf{x})\,, \tag{5}$$

，它保持我们数据集（ φ所诱导的标签（4）（1）。换言之，我们希望找到函数 $\mathbf{f}(\mathbf{x})$，使得

$$\vec{\varphi}(\mathbf{x}_i') \ \equiv \ \vec{\varphi}(\mathbf{f}(\mathbf{x}_i)) = \vec{\varphi}(\mathbf{x}_i), \quad \forall i \ = \ 1, 2, \ldots, m\,. \tag{6}$$

通常，给定的数据集会呈现出多种不同的对称性，接下来可以对其进行考察和分类

从群论的角度来看。为此，需要关注无穷小对称变换，并研究相应的一组不同生成元 $\{\mathbf{J}_\alpha\}$,$\alpha = 1,\ 2,\ \ldots\ldots,\ N_g$。如果满足闭合条件，即所有李括号 $[\,.\,,\,.\,]$ 都可以表示为该集合中已有生成元的线性组合，那么给定的一组生成元 $\{\mathbf{J}_\alpha\}$ 就构成一个李代数：

$$\big[\mathbf{J}_\alpha, \mathbf{J}_\beta\big] = \sum_{\gamma=1}^{N_g} a_{[\alpha\beta]\gamma} \mathbf{J}_\gamma. \tag{7}$$

系数$a_{[\alpha\beta]\gamma}$是我们数据集中对称群的结构常数（方括号索引记法提醒读者，它们的前两个指标是反对称的：$a_{[\alpha\beta]\gamma} = -a_{[\beta\alpha]\gamma}$）。最简单的代数（称为阿贝尔代数）是所有生成元都彼此对易、因而结构常数均为零的代数。本文的两个玩具示例中将会遇到这类代数。

3. 方法

在我们的方法中，我们使用一个神经网络 $\mathcal{F}_\mathcal{W}$ 对函数 $\mathbf{f}$ 进行建模，该网络的输入层和输出层均包含 n 个神经元，对应于数据点$\mathbf{x}'$的 n 个变换后的特征。在训练过程中，可训练的网络参数（权重和偏置）统称为 $\mathcal{W}$。这些参数会不断变化并收敛到训练后网络 $\mathcal{F}_{\widehat{\mathcal{W}}}$参数所对应的训练后取值$\widehat{\ }\mathcal{W}$，也就是说，帽号表示训练所得的结果。为确保网络具备所需的性质，我们设计了一个包含以下元素的损失函数。

不变性。 为了确保在变换（5）下的不变性，我们在损失函数中加入以下均方误差（MSE）项 L：

$$L_{\text{inv}}(\mathcal{W}, \{\mathbf{x}_i; \vec{y}_i\}) = \frac{1}{m} \sum_{i=1}^{m} \left[\vec{\varphi}\left(\mathcal{F}_\mathcal{W}(\mathbf{x}_i)\right) - \vec{y}_i\right]^2 . \tag{8}$$

使用该损失函数训练的神经网络将找到一个任意一般的（有限）对称变换 $\mathcal{F}_{\widehat{\mathcal{W}}}$，其参数由训练后的网络参数 $\mathcal{W}$的值决定$\widehat{\ }$。 $\mathcal{F}_{\widehat{\mathcal{W}}}$的具体实例取决于网络参数的初始化，因此，通过使用不同的初始化重复该过程，可以得到一系列对称变换。

无穷小性。 为了聚焦于对称生成元，我们将范围限制在恒等变换$\mathbf{I}$附近的无穷小变换$\delta\mathcal{F}$：

$$\delta\mathcal{F} \ \equiv \ \mathbf{I} + \varepsilon\, \mathcal{G}_\mathcal{W}\,, \tag{9}$$

其中 ε 是一个无穷小参数，而新神经网络 $\mathcal{G}$ 的参数 $\mathcal{W}$将被强制设为有限值。该

损失函数（8）随后可改写为

$$L_{\text{inv}}(\mathcal{W},\{\mathbf{x}_i;\vec{y}_i\}) = \frac{1}{m\varepsilon^2}\sum_{i=1}^{m}\left[\vec{\varphi}(\mathbf{x}_i+\varepsilon\mathcal{G}_{\mathcal{W}}(\mathbf{x}_i))-\vec{y}_i\right]^2, \tag{10}$$

分母中额外乘以一个因子 ε^2，以补偿一般变换的缩放比例为 ε (Craven 等人，2022年) 这一事实。此外，我们加入一个归一化损失项

$$L_{\text{norm}}(\mathcal{W},\{\mathbf{x}_i\}) = \tfrac{1}{m}\textstyle\sum_{i=1}^{m}\left[\|\mathcal{G}_{\mathcal{W}}(\mathbf{x}_i)\|-1\right]^2 + \tfrac{1}{m}\textstyle\sum_{i=1}^{m}\left[\|\mathcal{G}_{\mathcal{W}}(\mathbf{x}_i)\|-\overline{\|\mathcal{G}_{\mathcal{W}}(\mathbf{x}_i)\|}\right]^2. \tag{11}$$

最小化损失函数后，训练得到的神经网络 $\mathcal{G}_{\widehat{\mathcal{W}}}$将表示一个相应的生成元

$$\mathbf{J} \;=\; \mathcal{G}_{\widehat{\mathcal{W}}}, \tag{12}$$

其中

$$\widehat{\mathcal{W}} \equiv \arg\min_{\mathcal{W}}\left(L_{\text{inf}} + h_{\text{norm}}L_{\text{norm}}\right) \tag{13}$$

是学习得到的神经网络参数取值（h_{norm} 是通常取 1 的超参数）。在不同初始条件 $\mathcal{W}_0$下重复训练 N_g 次，可得到一组 N_g （通常彼此不同的）生成元 $\{\mathbf{J}_\alpha\}$、$\alpha = 1$、2、.....、N_g。

正交性。为确保生成元 $\{\mathbf{J}_\alpha\}$ 彼此不同，我们在损失函数中引入额外的正交性项

$$L_{\text{ortho}}(\mathcal{W},\{\mathbf{x}_i\}) = \frac{1}{m}\sum_{i=1}^{m}\sum_{\alpha<\beta}^{N_g}\left[\mathcal{G}_{\mathcal{W}_\alpha}(\mathbf{x}_i)\cdot\mathcal{G}_{\mathcal{W}_\beta}(\mathbf{x}_i)\right]^2. \tag{14}$$

群结构。为了检验某一组不同的生成元 $\{\mathbf{J}_\alpha\}$ 在前面步骤中找到的生成了一个群，我们需要检查代数的封闭性（7），例如通过最小化

$$L_{\text{closure}}(a_{[\alpha\beta]\gamma}) = \sum_{\alpha<\beta}\text{Tr}\left(\mathbf{C}^T_{[\alpha\beta]}\mathbf{C}_{[\alpha\beta]}\right), \tag{15}$$

关于候选结构常数参数$a_{[\alpha\beta]\gamma}$，其中封闭性失配定义为

$$\mathbf{C}_{[\alpha\beta]}(a_{[\alpha\beta]\gamma}) \equiv \left[\mathbf{J}_\alpha,\mathbf{J}_\beta\right] - \sum_{\gamma=1}^{N_g}a_{[\alpha\beta]\gamma}\mathbf{J}_\gamma. \tag{16}$$

由于 L_{closure} 是半正定的， $L_{\text{closure}} = 0$将表明该代数是封闭的，因此我们处理的是一个真正的（子）群。

原则上，生成元的数量 N_g 是一个必须预先指定的超参数。因此，当找到给定 N_g 值对应的闭合代数后，只能保证它是完整对称群的一个子代数，因此还必须继续尝试更高的 N_g 值。完整代数则对应于能够找到由不同生成元构成的闭合代数时， N_g 所能取的最大值。

4. 仿真设置

为了说明我们的方法，我们使用标准 MNIST 数据集（LeCun 和 Cortes，2010年），该数据集由 60000 张 28×28 像素的图像组成。换言之，该数据集（1）包含 $m = 60000$ 个样本，每个样本有 $n = 28\times 28 = 784$ 个特征。图像的标签是 0 到 9 的数字。该数据集的经典分类任务是构建一个预言机 $\vec{y} = \vec{\varphi}(\mathbf{x})$，它会计算输出层中的一个十维“logit”向量（这种原始输出通常会通过 softmax 函数归一化为相应的概率）。随后，图像被分类为$\text{argmax}(\vec{\varphi}(\mathbf{x}))$。

将其应用于 MNIST 数据集时，第2 节中的一般问题可以表述为：可以对数据集中的5）所示的变换类型应用于所有原始图像，使得 10 维 logit 向量$\vec{\varphi}(\mathbf{x})$保持不变，即满足对称方程（6）。请注意，如公式（6）所示，保持预言机函数的全部十个分量不变，是比仅要求每幅图像的预测结果保持不变强得多的条件——对于后者，只需确保映射$\mathbf{f}$ 满足以下条件

$$\arg\max(\vec{\varphi}(\mathbf{f}(\mathbf{x}))) = \arg\max(\vec{\varphi}(\mathbf{x})). \tag{17}$$

如第3节所述，并将在下一节的示例中加以说明，我们应用深度学习来推导所需的变换$\mathbf{f}$。该神经网络 $\mathcal{F}_{\mathcal{W}}$ 在 PyTorch (Paszke 等人，2019 年）中实现为顺序前馈神经网络。优化采用 Adam 优化器，学习率介于 3×10^{-5}和0.03之间。损失函数经过设计，旨在实现快速、高效的训练过程，无需大量超参数调优。训练和测试数据（1）来自标准 MNIST 数据集（LeCun 和 Cortes，2010 年）。

4.1. 从可忽略特征产生的平凡对称性

在继续之前，让我们通过讨论数据集中存在的平凡对称性来建立一些直觉，这类似于经典力学中对循环坐标（有时也称为可忽略坐标）的处理（Goldstein，1980）。在经典力学中，如果一个系统的拉格朗日量不包含某个坐标，那么该坐标就被称为循环坐标或可忽略坐标，其对应的共轭动量守恒。让我们来探讨一下，在我们的情形中是否存在这样的可忽略特征。

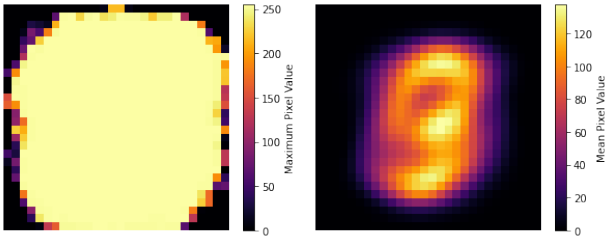

图 *1*。MNIST 数据集中每个像素的最大值（左图）或平均值（右图）的热力图。数据集中各个像素的取值范围为 0 到 255。

图 1 中的两张热力图分别显示了在整个数据集上取平均时每个像素的最大值（左图）和每个像素的平均值（右图）。图 1显示，图像的角落和边缘附近有许多像素完全不包含数据。这表明，一个稳健的分类器将不会受这些像素取值的影响，即这些像素的行为类似于经典力学中的可忽略坐标。另一个证据是，在所有特征中最高的、也就是信息量最少的主成分（此处未显示）是这些角落和/或边缘像素的线性组合。我们对这类平凡对称性不感兴趣。使用深度神经网络可能可以直接在图像之间找到非平凡映射，但训练的前提是存在一种能够捕捉数据集中非平凡结构的度量。[1] 这种度量的一个候选是推土机距离，它能够解决两个分布之间的最优传输问题；但正因为如此，其计算成本过高，无法用于损失函数。在下文中，我们将首先降低数据集的维度，并在相应的 ℓ-维潜在空间 $\mathbb{R}^\ell$中探索对称变换，其中 $\ell \ll n$。这是因为在潜在空间中，欧几里得度量能够捕捉丰富的结构。

4.2. 降维

如图 2 所示，我们不再在原始特征空间中寻找变换$\mathbf{f}$ $\mathbb{R}^n$，即保持预言机 $\vec{\varphi}$ （4）不变的对称变换，而是选择寻找变换 $\mathbf{g}$，它作用于潜在空间 $\mathbb{R}^\ell$，是保持诱导的预言机 $\vec{\psi}$： $\mathbb{R}^\ell \to \mathbb{R}^k$。为此，我们训练了一个自编码器，其架构

[1]例如，轻微旋转图像可能是相对于分类器的一种对称变换，但这会改变许多像素的值。因此，使用均方误差（MSE）这类朴素度量来衡量这两幅图像，所得距离会很大；而使用推土机距离这类更复杂的度量，所得距离则会很小。

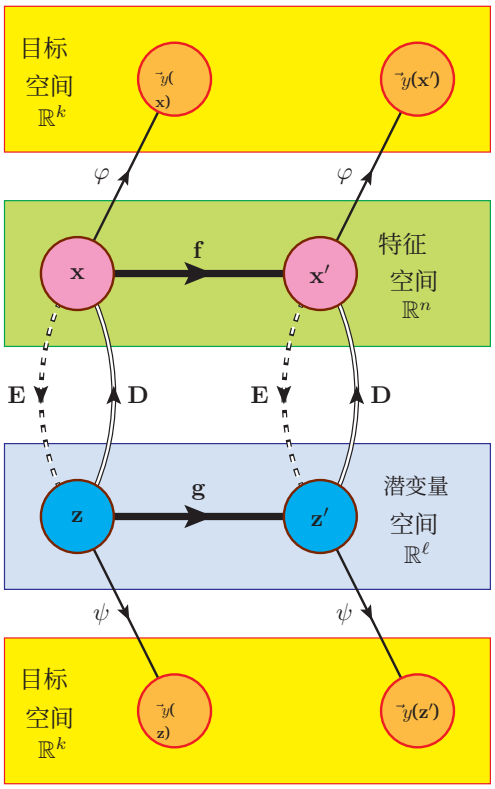

图 *2*。 本文讨论的不同数据处理步骤的流程图：编码器 $\mathbf{E}$：$\mathbb{R}^n \to \mathbb{R}^\ell$；解码器 $\mathbf{D}$： $\mathbb{R}^\ell \to \mathbb{R}^n$；特征空间中的变换 $\mathbf{f}$：$\mathbb{R}^n \to \mathbb{R}^n$；潜在空间中的变换 $\mathbf{g}$： $\mathbb{R}^\ell \to \mathbb{R}^\ell$；经过训练的向量预言机 $\vec{\varphi}$： $\mathbb{R}^n \to \mathbb{R}^k$，以及经过训练的向量预言机 $\vec{\psi}$：$\mathbb{R}^\ell \to \mathbb{R}^k$。如果 $\vec{y}(\mathbf{x}') = \vec{y}(\mathbf{x})$ 且 $\vec{y}(\mathbf{z}') = \vec{y}(\mathbf{z})$，那么变换 $\mathbf{f}$ 和 $\mathbf{g}$ 分别是对称变换。

如图 3 所示——它由一个编码器$\mathbf{E}$： $\mathbb{R}^n \to \mathbb{R}^\ell$ 和一个解码器 $\mathbf{D}$： $\mathbb{R}^\ell \to \mathbb{R}^n$。潜变量向量 $\mathbf{z}_i \in \mathbb{R}^\ell$定义如下

$$\mathbf{z}_i \equiv \mathbf{E}(\mathbf{x}_i), \qquad i = 1, 2, \ldots, m. \tag{18}$$

诱导出的分类器 $\vec{\psi}$： $\mathbb{R}^\ell \to \mathbb{R}^k$ 训练方式如下

$$\vec{y}_i = \vec{\psi}(\mathbf{z}_i) = \vec{\psi}(\mathbf{E}(\mathbf{x}_i)), \quad i = 1, 2, \ldots, m, \tag{19}$$

该分类器由一个包含三个隐藏层的全连接稠密神经网络构成，隐藏层大小分别为 128、128 和 32，并使用分类交叉熵损失进行训练。

有了这些工具，我们将寻找潜在空间中的对称变换，类比于公式（$\mathbf{g}$： $\mathbb{R}^\ell \to \mathbb{R}^\ell$ （9）

$$\mathbf{z}' = \mathbf{z} + \varepsilon\, \mathbf{g}(\mathbf{z}), \tag{20}$$

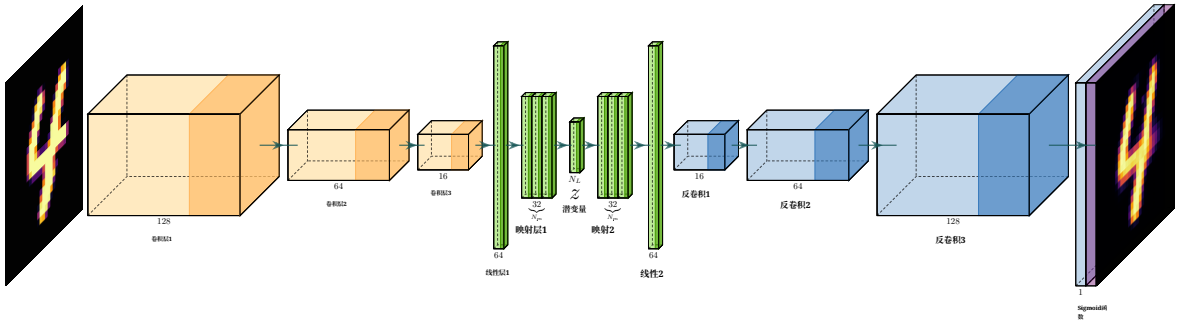

图 *3.* 我们的自编码器网络架构由编码器 **E** 和解码器 **D**组成。黄色模块为卷积层，绿色模块为全连接层，蓝色模块为转置卷积层。深色阴影区域表示 ReLU激活函数。

这些变换保持新的预言机不变 ψ，类比于公式（6）

$$\vec{\psi}(\mathbf{z}_i') \;\equiv\; \vec{\psi}(\mathbf{z}+\varepsilon\,\mathbf{g}(\mathbf{z})) = \vec{\psi}(\mathbf{z}_i), \quad \forall i \;=\; 1,2,\ldots,m\,. \tag{21}$$

单个变换 $\mathbf{g}$ 将由一个具有不同架构的全连接稠密神经网络表示，具体架构视练习的复杂程度而定，并使用第 3 节中描述的损失函数进行训练。一旦在潜在空间中找到这样的对称变换 $\mathbf{g}$，就可以借助解码器 **D** 展示并分析它对实际图像的影响。

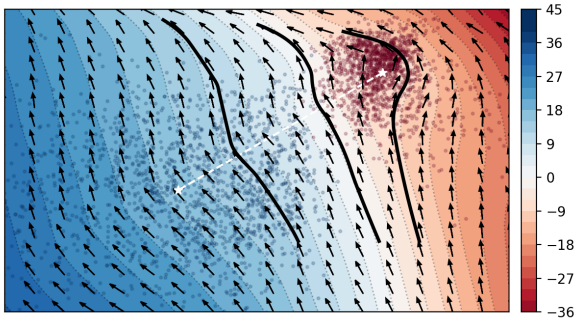

图*4*。 第 5.1 节中的练习结果以二维潜在空间呈现。红点和蓝点分别表示真实标签为 0 和 1 的验证图像，白色星形标记（由一条虚线连接）表示这两个簇的中心。热力图显示了 ψ 分类器输出层中单个神经元未经缩放的输出。叠加的向量场展示了我们的方法所发现的对称变换 $\mathbf{g}$。黑色实线是三条具有代表性的对称流线，用于图 5中的示意图。

5. 示例

对于下文第 5.3 节介绍的主要练习，我们将使用完整的数字集合和一个 16 维潜在空间；不过，在这个空间中，结果将难以绘图和可视化。因此，我们先从几个玩具示例入手，在这些示例中，我们只考虑两个类别，即数字零和数字一，并分别使用二维潜在空间（第 5.1 节）或三维潜在空间（第 5.2 节）。

5.1. 两个类别和 $\ell = 2$ 个潜变量

在本小节中，我们将在 $\ell = 2$ 个潜空间维度中进行一项简单的二元分类练习。我们只保留类别“0”和“1”的图像，然后按 3:1 的比例随机划分训练集和测试集。训练集用于训练图 3 中的自编码器和分类器 ψ。由于这是一个简单的二元分类示例，ψ 被设置为只有一个神经元的输出层，其逻辑值（原始输出）输入到 Sigmoid 函数中。结果如图 4 所示，绘制在两个潜变量构成的平面 $(z^{(1)}, z^{(2)})$中。红点和蓝点分别表示真实标签为 0 和 1 的验证图像集合，地。白色星形符号标出了这两个簇的中心，我们将它们称为数字[2] “柏拉图式的”零和一的图像。沿着两者之间的白色直虚线移动，我们会经过潜在空间中的一些点，这些点（经过解码后）会生成在柏拉图式的“0”和柏拉图式的“1”之间平滑插值的图像。图5的上方面板展示了这种变化。因此，沿白色虚线移动并非对称变换，因为它会改变图像的含义——

[2]从这个意义上说，它们是各自类别的理想代表。

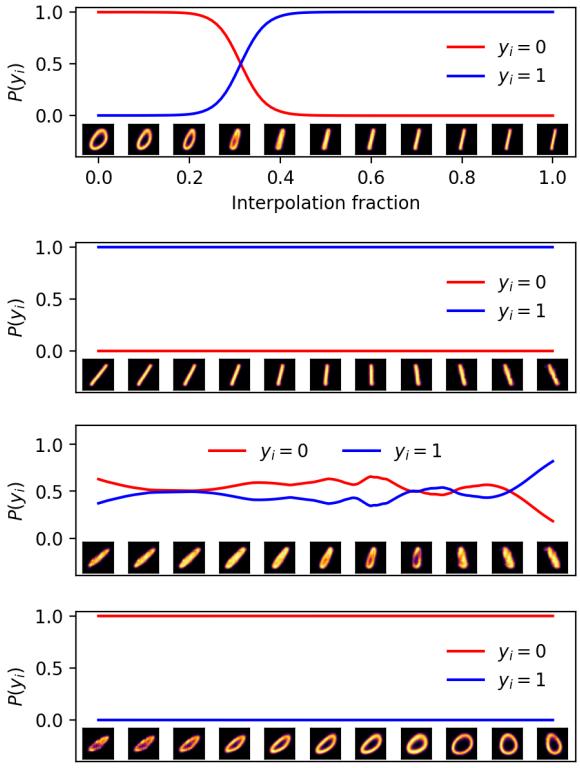

图*5*。通过沿着白色虚线（上图）、左侧黑色实线（第二幅图）、中间黑色实线（第三幅图）或右侧黑色实线（第四幅图），探索图4 中的潜在空间。每幅图都展示了沿相应潜在空间轨迹，图像属于零（红线）或一（蓝线）的似然。在每幅图的底部，我们展示一排应用解码器 $\mathbf{D}(\mathbf{z})$ 后得到的代表性图像。

义。我们关注的是沿正交方向的移动，因为在这个方向上，尽管实际图像发生变化，分类器对它的解读却保持不变。

我们训练好用于$\mathbf{g}$的神经网络后，便会在潜在空间中得到一个向量场。该向量场的流由图4中的黑色箭头表示。三条黑色实线是三条具有代表性的对称流线，用于说明图5底部三个面板中的内容。最左侧的流线经过接近柏拉图式数字一的位置，因此代表一系列被分类器以极高概率判定为“数字一”的图像（见图 ψ 5第二个面板中的蓝线）。第二个面板中对应的一行解码图像显示，对称变换的效果是使数字“1”逆时针旋转。

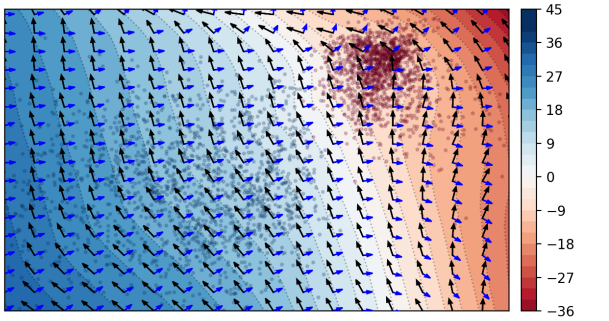

图 *6*。 与图 4相同，但要求两个相互正交的对称变换 $\mathbf{g}_1$ （黑色箭头）和 $\mathbf{g}_2$ （蓝色箭头）。

另一方面，图4中最右侧的流线穿过接近柏拉图式“数字零”的区域，因此其中的图像几乎必然会被分类器判定为数字零，这一点可由图5最后一个面板中的红线看出。最后一个面板中对应的一行解码图像显示，对称变换不仅会使数字“0”逆时针旋转，还会同时拉伸并放大图像。

图 4 中的中间流线穿过两个类别簇之间的边界区域，因此会生成分类器难以明确判别的图像。图5的第三幅图证实了这一点，其中显示沿该流线，“零”和“一”的似然相近[3] 。解码图像确实令人困惑，连人类也难以解读；它们也像其他图中的图像一样，以类似方式逆时针旋转。

我们方案的一个重要特点是能够找到多个非平凡对称性，反之，也能确定何时不可能存在其他对称性。由于我们选择的压缩表示只有两个维度，其中一个维度受预言机约束，因此只剩下一个对称自由度。因此，如果我们尝试寻找第二个对称性（如第 **3** 节所述，同时训练两个神经网络来学习 g），训练将收敛到一个较大的损失值，这意味着无法同时满足所有期望条件。具体而言，不变性要求潜在流沿着等似然度曲线流动，从而迫使它们彼此对齐，这与正交条件相矛盾。这

[3]值得注意的是，这条直线有些不稳定——这是因为该区域中的样本非常少，因此生成元模型在这里不太可靠。

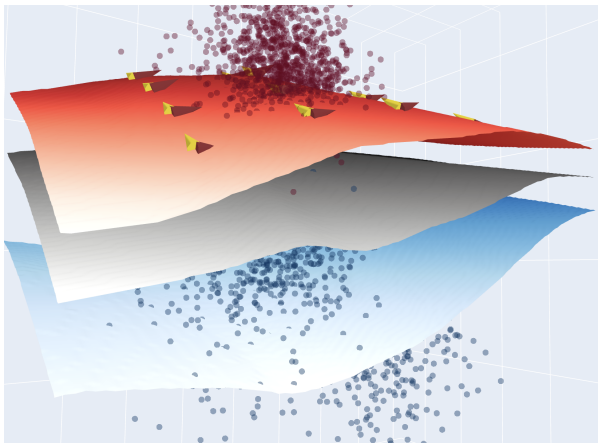

图 7。第 5.2 节中示例的三维潜在空间示意图。红色（蓝色）点分别表示真实标签为 0 和 1 的验证图像。三个曲面显示预言机的三个等值集 ψ，其“零”的似然分别为 0.9999、0.5 和 0.0001（从上到下）。黄色和红色箭头表示通过我们的方法找到的两个潜在流中采样的向量（为简单起见，我们只在上方曲面上采样点）。

一困境在图 6 中有所展示；我们针对 $N_g = 2$ 的情况重复了前面的练习。分别用黑色和蓝色箭头表示的两条潜流，试图在这些相互冲突且无法调和的要求之间取得最佳平衡。

5.2. 两个类别与 $\ell = 3$ 个潜变量

在本小节中，我们将玩具示例的潜在空间扩展到 $\ell = 3$ 个维度。与前一小节中的示例不同，这将使我们能够找到第二个非平凡的潜在流，它可以生成一种对称性。分析过程与之前相同，只是重新训练了带有三维瓶颈层的自编码器。最终结果见图 7，其中展示了三维潜在空间，以及形成两个簇的验证数据，分别对应数字零和数字一。随后，我们叠加了三个曲面，它们是预言机的等值集 ψ，其“零”的似然分别为 0.9999、0.5 和 0.0001。一般的局部对称变换与这些曲面相切，这恰好也是我们训练神经网络以寻找单个对称变换g。

我们的方法真正强大之处在于能够同时找到多个正交对称生成元。事实上，当我们针对 $N_g = 2$的情况重新训练时，能够找到两个正交向量流，如图7 中的黄色和红色箭头所示（为简便起见，我们仅展示从顶面采样的少量向量）。请注意，所有箭头都与该等值面相切，

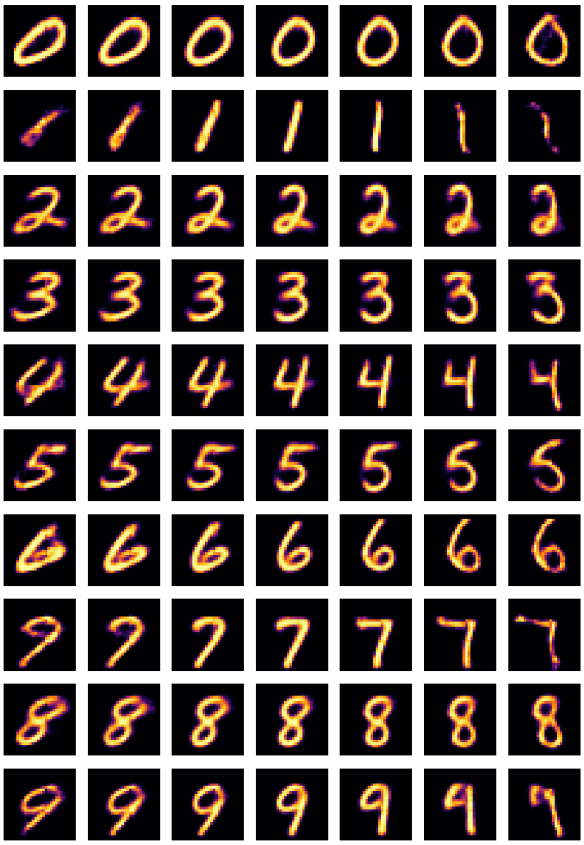

图 8。 沿第 5.3 节中发现的 16 维潜在流的流线对图像进行对称形变。中间列的图像表示数据集中的“柏拉图式”数字（十个类别各自的簇中心）。每一行中的其余六幅图像，均是沿着或逆着经过相应柏拉图式图像的流线移动得到的。

这正是不变潜流所预期的。

我们选择这个特定的玩具示例，是因为它代表了存在多个非平凡生成元、同时又能够轻松可视化的最复杂情形。在下一小节推广到更高维度后，这样简单的可视化将不再可能，但结果仍然可以用类似的方式直观地理解。

5.3. 十个类别和 $\ell = 16$潜变量

在本小节中，我们给出最后一个示例，其中考虑数字的全部十个类别，并使用带有 $\ell = 16$维度瓶颈层的自编码器。为了提升性能，我们发现，将映射层（即紧邻瓶颈层之前和之后的映射层）数量从 1 增至 3 很有帮助。与之前一样，我们训练了一个具有 10 个 Softmax 输出的分类器 $\vec{\psi}$以及一个用于对称变换的神经网络 g。结果如图 8所示，其中

与图 5 类似，我们展示了沿潜在流流线排列的一系列解码图像。图 8中央的图像表示数据集中的柏拉图数字。我们从这十个起点分别沿对应的流线前进 $\varepsilon = \pm 10^{-3}$ 步。中央图像左右两侧各三幅图像，分别是在每个方向上经过 2000、4000 和 6000 步后得到的。每一行中的图像都被正确分类，预测概率为 1.0。这验证了我们的方法适用于单个对称流的情况。

我们还同时训练了多个生成元，并发现即使生成元数量相对较多（$N_g \sim 10$），也存在非平凡的正交流，这暗示着非阿贝尔对称性的存在。本研究只是初步探索这一前景广阔的新研究方向，旨在揭示复杂数据集丰富的对称结构；相关工作留待未来开展。最终目标应是透彻理解其代数和拓扑性质。

6. 总结与结论

本文研究了数据科学中一个基础问题，该问题在许多领域都很常见：带标签数据集的对称性是什么？又该如何识别其群结构？为此，我们采用了一种深度学习方法，利用全连接神经网络对一般的对称变换及其对应的生成元进行建模。我们使用专门构造的损失函数训练网络，以确保所需的对称性，即：i）变换后的样本保留预言机输出（不变性）；ii）变换并非平凡变换（归一化）；iii）存在多个对称性时，学习到的变换必须彼此不同（正交性）；iv）这些变换必须构成一个代数（封闭性）。

我们利用了这样一个事实：非平凡对称性会在自动编码器学习到的潜在空间中诱导出相应的流。我们方法的一个重要优势在于，不需要预先了解数据中可能存在的对称性；也就是说，我们不是从预定义的可能性列表中检验对称性，而是直接学习对称性。

我们在标准 MNIST 手写数字数据集上展示了我们方法的性能，结果表明，学习到的变换对应于图像中具有可解释性的非平凡变换。未来的工作可以将这一方法扩展到更广泛的数据和对称性类型。

参考文献

Barenboim, G.、Hirn, J. 和 Sanz, V. 对称性遇见人工智能。*SciPostPhys.*，11:014，2021年。doi: 10.21468/SciPostPhys.11.1.014.

Bogatskiy, A.、Anderson, B.、Offermann, J. T.、Roussi, M.、Miller, D. W. 和 Kondor, R. 用于粒子物理的LorentzGroup等变神经网络。2020年6月。

Bogatskiy, A. 等。面向物理学的对称群等变架构。载于*2022*年*Snowmass*夏季研究，2022年3月。

Butter, A.、Kasieczka, G.、Plehn, T. 和 Russell, M. 基于深度学习的顶夸克标记与洛伦兹层。*SciPost Phys.*，5(3):028，2018年。doi：10.21468/SciPostPhys.5.3.028。

Carifio, J., Halverson, J., Krioukov, D., and Nelson, B. D.《弦景观中的机器学习》。《*JHEP*》，09：157，2017年。doi：10.1007/JHEP09(2017)157。

Chen, H.-Y., He, Y.-H., Lal, S., 和 Zaz, M. Z.《共形场论中的机器学习练习曲》。6 2020年。Chen, H.-Y., He, Y.-H., Lal, S., 和 Majumder.《机器学习李结构及其在物理学中的应用》。物理评论快报*B*辑，817:136297，2021年。doi: 10.1016/j.physletb.2021. 136297。Cranmer, M., Greydanus, S., Hoyer, S., Battaglia, P., Spergel, D., 和 Ho, S.《拉格朗日神经网络》，2020年。Craven, S., Croon, D., Cutting, D., 和 Houtz.《机器学习流形》。物理评论*D*辑，105(9):096030，2022年。doi: 10.1103/PhysRevD.105.096030。Desai, K., Nachman, B., 和 Thaler, J.《通过深度学习发现对称性》。物理评论*D*辑，105(9):096031，2022年。doi: 10.1103/PhysRevD.105.096031。Dillon, B. M., Kasieczka, G., Olischlager, H., Plehn, T., Sorrenson, P., 和 Vogel, L.《对称性、安全性与自监督》。*SciPostPhys.*，12(6):188，2022年。doi: 10.21468/SciPostPhys.12.6.188。Fenton, M. J., Shmakov, A., Ho, T.-W., Hsu, S.-C., White- son, D., 和 Baldi, P.《采用保持对称性的注意力网络进行无排列的多喷注事例重建》。物理评论*D*辑，105(11):112008，2022年。doi: 10.1103/PhysRevD.105.112008。Forestano, R. T., Matchev, K. T., Matcheva, K., Roman, A., Unlu, E., 和 Verner, S.《从第一性原理出发的深度学习对称性及其李群、代数与子代数》。*arXiv* 电子印本，文章编号 arXiv:2301.05638，2023年1月。doi: 10.48550/arXiv.2301.05638。Goldstein, H.《经典力学》。Addison-Wesley，1980年。

Gong, S., Meng, Q., Zhang, J., Qu, H., Li, C., Qian, S., Du, W., Ma, Z.-M., 和 Liu, T.-Y.《用于喷注标记的高效洛伦兹等变图神经网络》。*JHEP*，07:030，2022年。doi: 10.1007/JHEP07(2022)030。

Gross, D. J. 对称性在基础物理学中的作用。美国国家科学院院刊，93(25):14256–14259，1996年12月。doi: 10.1073/pnas.93.25.14256。

Gruver, N., Finzi, M., Goldblum, M. 和Wilson, A. G. 用于衡量学习到的等变性的李导数，2022年。

Hao, Z., Kansal, R., Duarte, J. 和 Chernyavskaya, N. 洛伦兹群等变自编码器。*arXiv* 电子预印本，文章编号 arXiv:2212.07347, 2022年12月。doi: 10.48550/a rXiv.2212.07347。

He, Y.-H. 机器学习弦景观。物理评论快报*B*辑，774:564–568，2017年。doi: 10.1016/j.physletb.2017.10 .024。

Iten, R., Metger, T., Wilming, H., del Rio, L. 和 Renner, R. 使用神经网络发现物理概念。物理评论快报, 124(1):010508, 2020年1月。doi: 10.1103/PhysRevLett.124.010508。

Kanwar, G., Albergo, M. S., Boyda, D., Cranmer, K., Hackett, D. C., Racani`ere, S., Rezende, D. J. 和 Shanahan, P. E. 用于格点规范理论的等变流采样。物理评论快报, 125(12):121601, 2020年。doi: 10.1103/PhysRevLett.125.121601。

Krippendorf, S. 和 Syvaeri, M. 使用神经网络检测对称性。*arXiv* 电子预印本, 文献号 arXiv:2003.13679, 2020年3月。doi: 10.48550/arXiv.2003.13679。

LeCun, Y. 和 Cortes, C. MNIST手写数字数据库。2010年。

Li, C.、Qu, H.、Qian, S.、Meng, Q.、Gong, S.、Zhang, J.、Liu, T.-Y. 和 Li, Q. 洛伦兹对称设计能否提升喷注物理中的网络性能？*arXiv* 电子预印本，文章编号 arXiv:2208.07814, 2022年8月。doi：10.48550/arXiv.2 208.07814。

Liu, Z. 和 Tegmark, M. 从轨迹中学习机器学习守恒定律。物理评论快报，126(18):180604，2021年。doi：10.1103/PhysRevLett.126.180604。

Liu, Z. 和 Tegmark, M. 机器学习隐藏对称性。物理评论快报，128(18):180201，2022年。doi：10.1103/PhysRevLett.128.180201。

Moskalev, A., Sepliarskaia, A., Sosnovik, I., 和 Smeul- ders, A. LieGG：研究学习得到的李群生成元。*arXiv* 电子预印本，文章编号 arXiv:2210.04345，2022年10月。doi：10.48550/arXiv.2210.04345。

Noether, E. 不变变分问题。哥廷根科学学会消息，*G¨ottingen* 数学物理学部，1918年：235–257，1918年。

Paszke, A.、Gross, S.、Massa, F.、Lerer, A.、Bradbury, J.、Chanan, G.、Killeen, T.、Lin, Z.、Gimelshein, N.、Antiga, L.、Desmaison, A.、Kopf, A.、Yang, E.、DeVito, Z.、Raison, M.、Tejani, A.、Chilamkurthy, S.、Steiner, B.、Fang, L.、Bai, J. 和 Chintala, S. PYTORCH：一种命令式风格的高性能深度学习库。载于《神经信息处理系统进展第*32*卷》，第8024–8035页。Curran Associates, Inc.，2019年。

Ruehle, F. 弦理论中的数据科学应用。《物理报告》，839：1–117，2020年。doi：10.1016/j.physrep.2019.09.005。

Shmakov, A.、Fenton, M. J.、Ho, T.-W.、Hsu, S.-C.、Whiteson, D. 和 Baldi, P. SPANet：利用保持对称性的注意力机制，为粒子物理学实现广义的无排列集合分配。《*SciPostPhys.*》，12(5)：178，2022年。doi：10.21468/SciPostPhys.12.5.178。

Wetzel, S. J., Melko, R. G., Scott, J., Panju, M. 和 Ganesh, V. 通过解读孪生神经网络发现对称不变量和守恒量。物理评论研究，2(3):033499，2020年。doi: 10.1103/ physrevresearch.2.033499。

Wigner, E.、Griffin, J. 和 Griffin, J. 群论及其在原子光谱量子力学中的应用。《纯粹与应用物理》：专著和教科书系列。学术出版社，1959年。ISBN 9780127505503。

Wu, T. 和 Tegmark, M. 迈向用于无监督学习的人工智能物理学家。《物理评论*E*》，100(3)，2019年9月。doi: 10.1103/physreve.100.033311。