

什么是流形学习

什么是流形呢？让我们看图 1（左）所示的一个数据分布，可以看到虽然数据是三维的，但所有数据都分布在一个蛋卷形的二维曲面上。这个曲面称为一个二维流形。对于分布在这个流形上的数据来说，很重要的不是它在三维空间中的坐标，而是在这个二维流形上的位置。因此，给定一个三维数据集，我们希望把这个二维流形学习出来，并确定每个数据样本在这个二维流形上坐标，如图 1（右）所示。这就是流形学习。

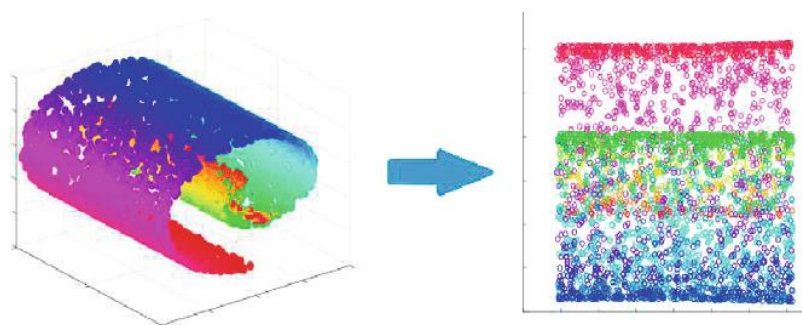


图 1: 流形学习的目的是学习高维空间中的低维曲面，将高维坐标映射为平面上的低维坐标。

流形学习有什么用处呢？这是因为在实际机器学习任务中，我们遇到的数据大部分都是高维的，直接建模会带来很大问题。以图片数据为例，MNIST 手写数据 $28 \times 28 = 784$ 维，imageNet 图片数据经过预处理后有 $256 \times 256 = 65,536$ 维。虽然维度高的离谱，但绝大部分数据只分布在一个低维流形上，如图 2 所示。如果我们能把数据映射到这个低维流形上，可以减小维度灾难的影响，降低建模的数据需求。另外，高维数据很不直观，如果我们可以把数据放在一个二维或三维流形中展示出来，可以很直观地了解数据的分布情况，有利于模型选择和结果分析。总结来说，降维和可视化，是流形学习的两个根本目的。

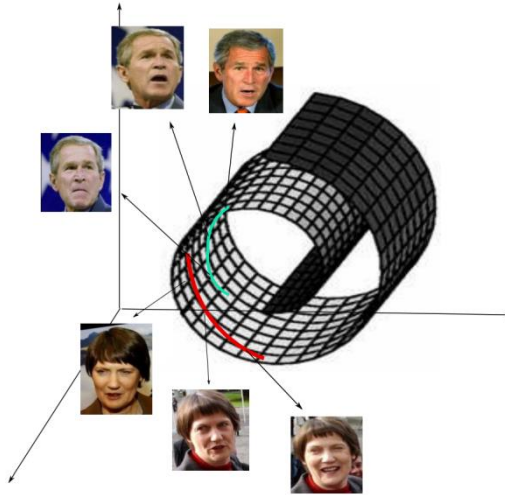


图 2：人脸图片数据的流形示意图[1]

流形学习是一种典型的无监督学习方法，不涉及数据标注问题，通过分析数据样本点之间的关系来发现低维流形面。主成分分析（Principle Component Analysis, PCA）或多维标度（Multi-Dimensional Scale, MDS）是典型的流形学习方法。这两种方法通过寻找一个线性映射将数据投影到一个低维平面上，如果数据集中在一个低维超平面上，则这种线性方法可以得到较好的学习效果。反之，如果数据集中在一个曲面上，则需要更复杂的非线性方法。常用非线性嵌入方法包括局部线性嵌入（Local Linear Embedding, LLE）、谱嵌入（Spectral Embedding）和 t-SNE（Student's t Stochastic Neighbour Embedding）等。不论是哪种方法，基本思路都是在低维空间中重构高维空间中数据样本间的相邻关系。换句话说，如果在原始空间中两个样本是相近的，那么他们在低维流形上也应该是相近的，反之亦然。

图 3 是采用各种流形学习方法对 MNIST 手写数字图片做二维嵌入得到的低维分布，其中每种颜色代表一个数字。可以看到，t-SNE 模型得到的二维流形对数字有更好的区分性，证明这一模型对数据有更强的低维表征能力。正是因为这个原因，t-SNE 被广泛应用于数据可视化中。

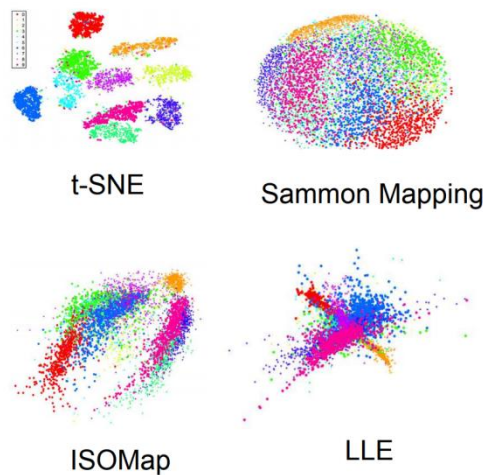


图 3：不同非线性流形学习方法对 MNIST 手写数字图片的二维嵌入结果[2]。

[1] 王东, 机器学习导论, 清华大学出版社, 2021.2.

[2] Maaten Lvd, Hinton G (2008) Visualizing data using t - SNE. Journal of Machine Learning Research 9:2579 - 2605