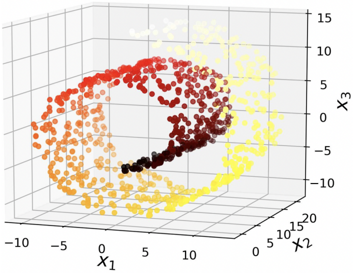
Dimensionality Reduction for Visualization

Manifold Learning

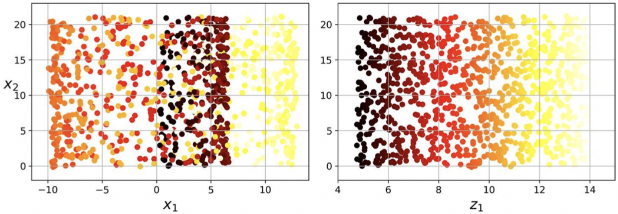
- A d-dimensional manifold is a portion of an n-dimensional space, $d < n$, that locally resembles d-dimensional Euclidean space
  - Locally resembles essentially means smooth in the neighborhood around any given point

- **Manifold hypothesis**: real-world, high-dimensional data often lie on low-dimensional manifolds embedded in the high-dimensional space

A Low-Dimensional Manifold in High-Dimensional Space



The Swiss roll dataset



Projecting (left panel) versus learning and unrolling the Swiss roll's manifold

t-Distributed Stochastic Neighbor Embedding[1]

- **t-Distributed Stochastic Neighbor Embedding** (t-SNE) maps high-dimensional instances to low-dimensional points (often 2D or 3D) and is often useful for visualization
    - Similar instances have points that are proximal
    - Dissimilar instances have points that are distant

- Algorithm
    1. Create a probability distribution over pairs of high-dimensional objects
    2. Define a probability distribution over low-dimensional points
    3. Select low-dimensional points so that the probability distribution from Step 1 is similar to that from Step 2

[1] For more information about t-SNE, see here

t-SNE Similarity

- Assume there are M high-dimensional objects
- Probability distribution for the high-dimensional objects $\mathbf{x}_i$ and $\mathbf{x}_j$

$$p_{j|i} = \frac{\exp\left(-\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2\right) / \left(2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\left\|\mathbf{x}_i - \mathbf{x}_k\right\|^2\right) / \left(2\sigma_i^2\right)}, \quad p_{ij} = \frac{p_{j|i} + p_{i|j}}{2M}$$

- Probability distribution for the low-dimensional points $\mathbf{y}_i$ and $\mathbf{y}_j$

$$q_{ij} = \frac{\left(1 + \left\|\mathbf{y}_i - \mathbf{y}_j\right\|^2\right)^{-1}}{\sum_k \sum_{l \neq k} \left(1 + \left\|\mathbf{y}_k - \mathbf{y}_l\right\|^2\right)^{-1}}$$

Relating the Distributions

- Select the low-dimensional points, $\mathbf{y}$, by minimizing the *Kullback–Leibler divergence*

$$KL\left(p \mid q\right) = -\sum_{i \neq j} p_{ij}\left(\log\left(q_{ij}\right) - \log\left(p_{ij}\right)\right) = \sum_{i \neq j} p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right)$$

- Note: KL divergence is based on the *entropy* of a probability distribution

$$H\left(p\right) = -\sum_{i} p_i \log\left(p_i\right)$$