Persistent Homology for the Evaluation of Dimensionality Reduction Schemes

Bastian Rieck    Heike Leitte

Interdisciplinary Center for Scientific Computing
Heidelberg University
Motivation
Motivation
Motivation
Which one is more suitable?
Contribution

Data Embeddings

Persistent homology
Data descriptors

Bastian Rieck
Evaluating Dimensionality Reduction Schemes
The plan

Compare embeddings with the original data using the topology of data descriptors.
Topology

Topology = connectivity information

- Connected components
- Tunnels
- Voids
- ...
Persistent homology

1-dimensional example
Persistent homology

- Framework for characterizing multivariate data sets.
- Summarize changes in connectivity as the scale of the object changes.
- Stable comparison of objects via their summaries ("fingerprint").
Comparing persistence diagrams

Bottleneck distance

\[ \|f - g\|_\infty \]

\[ W_\infty(X, Y) \]
Multivariate data

Neighbourhood graph
Data descriptor

\[ f(x) = \frac{-1}{k} \sqrt{\sum_{i=1}^{k} \text{dist}^2(x, n_i)} \]

- Use distance to \( k \) nearest neighbours to estimate density
- Stability under perturbations in the data
- No preference for any dimensionality reduction method
Workflow

Original data

Embeddings

Data descriptors

Nb'hood graph

Persistent homology

$W_\infty(\cdot,\cdot)$

$W_\infty(\cdot,\cdot)$
Quality_{Embedding} = W_{\infty}(D_{Original}, D_{Embedding})
Example
Swiss hole

- Original Data
- HLLE
- t-SNE
- SPE

<table>
<thead>
<tr>
<th>Method</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLLE</td>
<td>1.66</td>
</tr>
<tr>
<td>t-SNE</td>
<td>2.26</td>
</tr>
<tr>
<td>SPE</td>
<td>10.67</td>
</tr>
<tr>
<td>Isomap</td>
<td>2.26</td>
</tr>
<tr>
<td>PCA</td>
<td>3.42</td>
</tr>
</tbody>
</table>

Bastian Rieck
Evaluating Dimensionality Reduction Schemes
Global quality by calculating distances between embedding and original data.
Global quality by calculating distances between embedding and original data.

What about local quality?
Local quality information

Quality

- Good
- Medium
- Bad
Results

Isomap faces

Results

Isomap faces: Parameter study

- **MDS**: 1.65
- **t-SNE**: 2.52
- **Isomap** (k=8): 3.55
- **Isomap** (k=16): 3.62
- **Isomap** (k=10): 5.07

Bastian Rieck

Evaluating Dimensionality Reduction Schemes
Results

Climate data: Different aspects

- **SPE**: 3.12
- **PCA**: 3.14
- **RP**: 3.80
- **t-SNE**: 4.51
Summary

Novel evaluation scheme for dimensionality reduction:

- Calculate data descriptors on embeddings and on the original data.
- Use persistent homology to obtain a stable quantification of differences.
- Annotate scatterplots with local and global quality information.