We construct a Hierarchical Clustering of the set of items from pairwise distances.

The intrinsic dimension is commonly estimated using a covering of the data:

\[ m(r) = r^{-d} \]

Applications include: intrinsic dimension analysis include Internet topology analysis, computer vision, and computational finance.

This hierarchical clustering defines a nested covering of the objects. We annotate each cluster with the covering diameter.

By pruning this tree structure, we estimate the minimum covering, \( m(r) \), for a specified covering diameter, \( r \).

Using values of \( m(r) \) and \( r \), we estimate the intrinsic dimension.

Validation is performed with respect to fractals where the non-integer intrinsic dimension is known.

Consider pairwise distances that conform to a clustering structure.

Complete linkage condition:

Given three objects \( \{i, j, k\} \) where:

\[ \{i, j\} \in C \] and \( k \notin C \)

\[ d_{i,j} < \min(d_{i,k}, d_{j,k}) \]

If this condition holds, the estimated dimension using ClusterDimension converges to the true intrinsic dimension.

\[ \lim_{r \to 0} -\frac{\log m(r)}{\log r} = d \]

Intrinsic Dimension Estimation RMSE
(averaged across 10 realization of 750 sampled points)

<table>
<thead>
<tr>
<th>Dimension Estimation Method</th>
<th>Koch Curve</th>
<th>Sierpinski Triangle</th>
<th>Sierpinski Carpet</th>
<th>1-D Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClusterDimension</td>
<td>0.030</td>
<td>0.083</td>
<td>0.062</td>
<td>0.041</td>
</tr>
<tr>
<td>Maximum Likelihood [1]</td>
<td>0.062</td>
<td>0.065</td>
<td>0.245</td>
<td>0.043</td>
</tr>
<tr>
<td>Box Counting [2]</td>
<td>0.172</td>
<td>0.273</td>
<td>0.444</td>
<td>0.350</td>
</tr>
<tr>
<td>Correlation Dimension [3]</td>
<td>0.280</td>
<td>0.250</td>
<td>0.491</td>
<td>0.191</td>
</tr>
<tr>
<td>Minimum Spanning Tree [4]</td>
<td>0.248</td>
<td>0.221</td>
<td>0.091</td>
<td>0.131</td>
</tr>
<tr>
<td>PCA</td>
<td>0.738</td>
<td>0.416</td>
<td>0.107</td>
<td>0.0</td>
</tr>
</tbody>
</table>

References