Video

Bookmark this page

Isomap Gotchas!

that

the inner sample distances
are preserved as much as

your original dataset.

as

your original dataset.

0:55 / 0:55 1.25x HD

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Running Isomap is a lot slower than PCA since a lot more is happening under the hood, particularly for large `n_neighbors` values, but it provides a simple way to analyze and manipulate high dimensional samples in terms of its intrinsic nonlinear degrees of freedom. This is because isomap attempts to keep the global structure of your data as it reduces its dimensionality. You can use it in any case where nonlinear geometry degrades the effectiveness of PCA.

Isomap is also a bit more sensitive to noise than PCA. Noisy data can actually act as a conduit to short-circuit the nearest neighborhood map, cause isomap to prefer the 'noisy' but shorter path between samples that lie on the real geodesic surface of your data that would otherwise be well separated.

When using unsupervised dimensionality reduction techniques, be sure to use the feature scaling on all of your features because the nearest-neighbor search that Isomap bases your manifold on will do poorly if you don't, and PCA will prefer features with larger variances. As explained within the labs' source code, SciKit-Learn's StandardScaler is a good-fit for taking care of scaling your data before performing dimensionality reduction.
