

# ISOMAP

Ariah Klages-Mundt

**ISOMAP description** Multidimensional Scaling (MDS) embeds an  $n$ -dimensional dataset of  $m$  points into a  $d$ -dimensional space in a way that approximately preserves pointwise distances. The output is embedded coordinate vectors  $y^i$  for each original point  $0 \leq i \leq m$ . This is typically done using the Euclidean distance matrix  $D$  and minimizing a squared error loss function, in which case the solution can be expressed in terms of the eigendecomposition of a variant of the distance matrix:

$$\tau(D) = -\frac{1}{2}HSH,$$

where  $H = I - \frac{1}{n}11^T$  and  $S$  is the matrix of squared distances. In particular, let  $\lambda^p$  and  $v^p$  be the  $p$ th eigenvalue and corresponding eigenvector of  $\tau(D)$  and  $v_i^p$  be the  $i$ th component of the  $p$ th eigenvector. Then the  $p$ th component of the  $d$ -dimensional embedded coordinate vector  $y^i$  is

$$y_p^i = \sqrt{\lambda^p} v_i^p. \tag{1}$$

MDS may not perform well if Euclidean distance is a poor measure of dissimilarity between points as nonlinear structures are invisible to MDS. As detailed in [1], Isometric Feature Mapping (ISOMAP) instead uses geodesic distance between points along the data manifold to measure pointwise dissimilarity. This can more accurately capture neighborhood relationships when there is nonlinear structure to the data.

The (k-)ISOMAP algorithm takes input parameters  $k$  and  $d$  and executes the following steps:

1. Construct the undirected  $k$ -nearest neighbors (knn) graph over the data points, as determined using a distance measure (e.g., Euclidean), with edges weighted by that distance measure.
2. Compute the shortest path distance matrix  $D_G$  between points in the knn graph using, for example, Floyd's algorithm.
3. Construct the  $d$ -dimensional MDS embedding of  $\tau(D_G)$  as detailed above.

**Analyzing residuals** The residual variance is calculated as described in [1] as

$$1 - R^2(\hat{D}_M, D_Y),$$

where

- $\hat{D}_M$  is each algorithm's estimate of the manifold distance matrix (for ISOMAP this is  $D_G$ , for MDS this is Euclidean distance),
- $D_Y$  is the Euclidean distance matrix of points in the  $d$ -dimensional embedding output from each algorithm,
- $R$  is the linear correlation coefficient taken between the flattened arrays of  $\hat{D}_M$  and  $D_Y$ .

Note that comparing the residual variance of ISOMAP to MDS doesn't allow us to judge whether ISOMAP is better than MDS for a given dataset. This inherently depends on whether Euclidean distance is a good measure of dissimilarity between the data points, which is not captured in this

residual variance. Plotting residual variance vs.  $d$  only lets us visualize the added explanatory power of additional dimensions in the embedding.

Note that although  $\tau(D_G)$  from ISOMAP will always be symmetric, it need not be positive definite. In that case, some eigenvalues can be negative. Due to the square root in Equation 1, some coordinates can end up being imaginary. Customarily, such dimensions tend to be excluded from ISOMAP, and so the residual variance may not go to zero with added dimensions. Alternatively, taking the eigenvectors (unscaled by the eigenvalues) as the embedding directly can also be effective in practice.

## References

- [1] J.B. Tenenbaum, V. de Silva, and J.C. Langford. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):23192323.