Structural Intrinsic Dimensionality

Stephane Marchand-Maillet – Oscar Pedreira – Edgar Chavez

Premise: Local dimensionality M is estimated pointwise from the variation of local density:

- From data: $$\mathrm{GED}(x_i) = \frac{\log \frac{n_2}{n_1}}{\log \frac{r_2}{r_1}}$$
- From angles: $$\mathrm{P}(\theta) = \frac{\Gamma(\frac{M}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{M-1}{2})}\sin^{M-2}(\theta) \quad \text{and} \quad \mathrm{Var}(\cos(\theta)) = \frac{1}{M}$$
- Using the Kissing number: $$(1+O(1))\sqrt{\frac{3\pi}{8}}\log\left(\frac{3}{2\sqrt{2}}\right)M^{\frac{3}{2}}\left(\frac{2}{\sqrt{3}}\right)^{M} \leq \mathrm{Kiss}(M) \leq (1+O(1))\sqrt{\frac{\pi}{8}}M^{\frac{3}{2}}2^{\frac{M}{2}}$$

Result may be spatially unstable since no correlation between local estimates

Proposal: Structural Intrinsic Dimensionality

Local dimensionality estimates arise from the strcuture of the dataset, captured by a graph:

1. Build a graph (spanner) capturing local relationship
   - e.g: Half-space partitioning graph (HSP)
2. Diffuse dimensionality information over the graph
   - e.g: PageRank-like algorithm or convolution

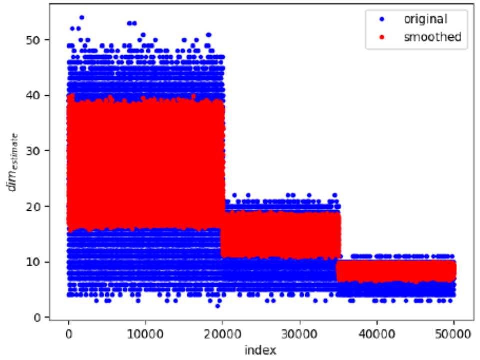
Outcome: a globally inferred local dimensionality estimate

Results: Structrural Regression

Regression of the initial dimensionality estimates (IID) over the HSP:

$$d_i^{(t+1)} \leftarrow d_i^{(t)} - \eta\left[(d_i^{(t)} - \epsilon_i) + \lambda\left(d_i^{(t)} - \sum_{x_j \in \mathcal{N}(x_i)} w_{ij} d_j^{(t)}\right)\right]$$

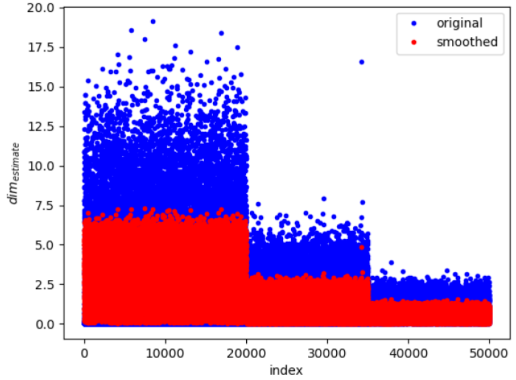
Variance of the estimate is better controlled



Results: Structrural Intrinsic Dimensionality

PageRank-like propagation of information: The dimensionality estimate emerges from the graph structure itself.

$$d_i^{(t+1)} \leftarrow \sum_{x_j \text{ s.t. } x_i \in \mathcal{N}(x_j)} w_{ji} d_j^{(t)}$$



Question:

What best graph structure to encode dimensionality?