Chapter 7 Manifold Forests

- Learning the structure of high-dimensional data & Mapping it onto a lower dimensional space
- Real data - large number of dimensions
- more direct visualization, easier data interpretation
- Properties of MF:
  1. computational efficiency
  2. automatic selection of discriminative features via information-based optimization
  3. other (i.e. code reusability)

7.1 Manifold Learning & Dimensionality Reduction in the lit

1. Principal Component Analysis (PCA) - linear
2. Isometric feature mapping

Laplacian eigenmaps - DR

→ try to preserve local pairwise point distances only

Medical Image Analysis

7.2 Specializing the Forest Model for Manifold L

Given a set of k unlabeled observations $\{v_1, v_2, \dots, v_i, \dots, v_k\}$ with $v_i \in \mathbb{R}^d$ we wish to find a smooth mapping $f: \mathbb{R}^d \to \mathbb{R}^{d'}$ with $f(v_i) = v_i'$ that approx. preserves the observations relative geodesic distances, with $d' \ll d$.

- collection of clustering trees
- Extra: a model of affinity bt. data points

7.2.1 The training objective function

RNO

$$\theta_j = \underset{\theta \in \mathcal{T}_j}{\arg\max}\, I(S_j, \theta)$$

$$I(S_j, \theta) = \log(|\Lambda(S_j)|) - \sum \frac{|S_j^i|}{|S_j|} \log(|\Lambda(S_j^i)|)$$

7.2.2 The Prediction Model

(7.3) $$P_t(v) = \frac{\pi_{\ell(v)}}{Z_t} \mathcal{N}(v; \mu_{\ell(v)}, \Lambda_{\ell(v)})$$

7.2.3 The Affinity Model

→ important to estimate similarity / affinity / distance between data points

Partition of input points:

$$\ell(v) : \mathbb{R}^d \to C \subset \mathbb{N}$$

Affinity matrix: $$W^T_{ij} = e^{-Q^t(v_i, v_j)}$$

Distance Q:

1. Mahalanobis affinity
2. Gaussian affinity
3. Binary Affinity ← simplest & interesting

$$Q^t(v_i, v_j) = \begin{cases} 0 \\ \infty \end{cases}$$

7.2.4 The Ensemble Model

→ smoother affinity matrix

$$W = \frac{1}{T} \sum_{t=1}^{T} W^T \quad (7.9)$$

→ add robustness

7.2.5 Estimating the Embedding Function

→ Laplacian eigenmaps

matrix: $L = I - \Upsilon^{-\frac{1}{2}} W \Upsilon^{-\frac{1}{2}}$ (7.10)

$$v_i' = f(v_i) = (E_{i1}, \ldots, E_{ij}, \ldots, E_{id'})^T$$

7.2.6 Mapping Previously unseen points

→ interpolate the point position given the already available embedding

$$v' = \frac{1}{T} \sum_t \frac{1}{\gamma^t} \sum \left(e^{-Q^t(v, v_i)} f(v_i) \right)$$

7.2.7 Properties and Advantages

(1) Ensemble Clustering for Distance Estimation

complex data → use pairwise affi. defined by tree structure

(2) Choosing the Feature Space

→ automatically select discriminative features and corresponding parameters.

(3) Computational Efficiency

① $d' << k$ bottom eigenvectors

② L is very sparse

(4) Estimating the Target Intrinsic Dimension-
-ality

Choose value d':

① profile of eigenvalues λ_j

② minimum number of eigenvalues

7.3 Experiments & Effect of Model Para-
-meters

(1) Forest size

Large T → positive effects (robustness to noise)

better approximate pairwise graph affinity

(2) Higher Dimensions

(3) Discovering the Manifold Intrinsic Dimensionality

small value of d'

Sharper Elbow:

→ higher values of T

→ Gaussian affinities