Classification by diffusion bases

Yosi Keller[1] Steven Damelin[2]

[1]School of Engineering, Bar-Ilan University

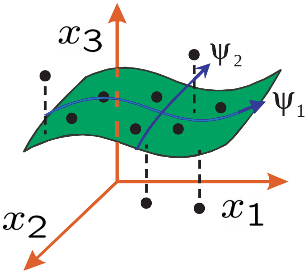
[2]Department of Mathematical Sciences, Georgia Southern University

Content

1. Spectral embedding vectors form adaptive basis
2. Density invariant embedding
3. One-Vs-All Classifier for high dimensional data
4. Expansion of the classification function using diffusion bases

Curse of dimensionality & Dimensionality Reduction

- Density estimation is difficult
- Computational cost of many algorithms grows exponentially with the dimension.
- Certain signals are in essence low-dimensional and their high dimensional representation is due to over sampling and noise.
- The high dimensional representation obscures the underlying low dimensional structure.
- Certain tasks only require low-dimensional representations.



Kernel methods

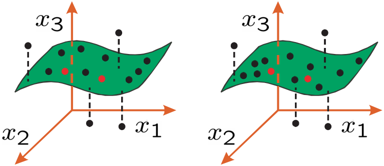
Definition

Given a dataset $\{x_i\}_{i=1..n}$:

1. Apply a p.s.d. kernel k to $\{x_i\}$. For instance: $w_{ij} = \exp(-\frac{d(x_i,x_j)}{\varepsilon})$, $\varepsilon > 0$.
2. Compute the eigenvectors of W: $w_{ij} = \sum_{l \geq 0} \lambda_l \psi_l(i) \psi_l(j)$,
3. The embedding is given by $\Psi(x_i) : x_i \mapsto \left(\lambda_1 \psi_1(x_i), \lambda_2 \psi_2(x_i), \ldots\right)$ $\|\Psi_t(x_i) - \Psi_t(z_i)\|^2_{L_2} = \sum_{l=0}^{n-1} \lambda_l^{2t} (\psi_l(x) - \psi_l(z))^2 =$ $w_{ii} + w_{jj} - 2w_{ij} = D(x,z)^2$

Density invariant embeddings I

[Coifman,Lafon,et. al,PNAS2005],[Keller,Lafon,Coifman,PAMI2006]



Same manifold, same points in $\mathbb{R}^3$, different densities

$\Downarrow$

different embeddings in $\mathbb{R}^2$

Density invariant embeddings II

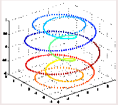
[Coifman,Lafon,et. al,PNAS2005],[Keller,Lafon,Coifman,PAMI2006]

Density invariant embedding

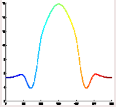
Given $w_{ij} = \exp(-\frac{\|x_i,x_j\|^2}{\varepsilon})$, define $q_i \triangleq \sum_j w_{ij}$,and form the new kernel $\widetilde{w}_{ij} = \frac{w_{ij}}{q_i q_j}$.

Now continue with the regular graph Laplacian using $\widetilde{w}_{ij}$.

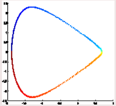
Curve



Density

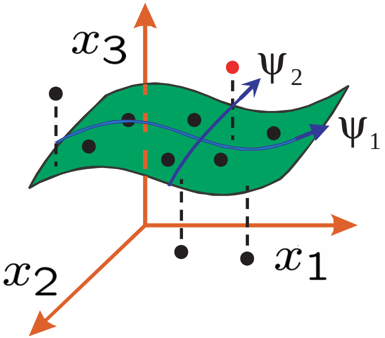


Graph Laplacian



Laplace-Beltrami

Out of sample extension I
[Keller,Lafon,Coifman,PAMI2006]



Given the parametrization $\{\psi_l\}$ computed using $N \gg 1$ samples, extend the embedding to the new point y, without re-embedding the $N+1$ data set.

- This is a multiscale Nyström extension.
- Differs from [Fowlkes, Belongie, Chung, Malik].

Out of sample extension II
[Keller,Lafon,Coifman,PAMI2006]

Spectral low pass extension

Given a p.s.d. symmetric $\widetilde{k}$ with $\widetilde{\varepsilon} \gg \varepsilon$ and its eigensystem $\left\{\widetilde{\psi}_l,\widetilde{\lambda}_l\right\}$. $\widetilde{\psi}_l$ can be approximated beyond $x \in \bar{X}$ by

$$\widetilde{\psi}_l(y) = \frac{1}{\lambda_l} \sum_{z \in X} \widetilde{k}(y,z)\widetilde{\psi}_l(z),\ \lambda_l > C,\ \forall y$$

and used to extend $\{\psi_l\}$

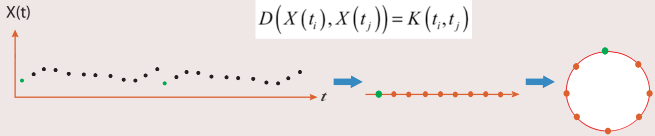
$$\psi_l = \sum_l \langle \psi_l, \widetilde{\psi}_l \rangle_X \widetilde{\psi}_l,\ \forall y$$

Spectral embedding vectors as an adaptive basis

The kernel k is p.s.d, hence its eigenvectors (the embedding) form an orthonormal system $\{\psi_i\}$.

Example

Given a periodic discrete signal $x(n)$, $n = 1 \ldots N$, and a time invariant metric $D(x(t_1), x(t_2)) = D(|t_1 - t_2|)$



- A circle is parameterized by $e^{\frac{2\pi i}{N} n}$, $n = 1 \ldots N$.
- For any kernel K, the Laplacian is a circulant matrix, diagonalized by the Fourier basis.

Sinc interpolation of periodic functions

- Start with the Fourier basis $\{\psi_i\}$ on the circle
  $f(x) = \sum_i a_i \psi_i(x)$, where $a_i = \langle f, \psi_i(x) \rangle$
- Extend $\{\psi_i\}$ from $X \to y$ using the Nystrom extension
  $\psi_i(y) = \frac{1}{\lambda_i} \sum_x K(x,y)\, \psi_i(x)$
- Extend f from $X \to y$
  $f(y) = \sum_i a_i \psi_i(y) = \sum_i a_i \frac{1}{\lambda_i} \sum_x K(x,y)\, \psi_i(x) =$
  $\sum_x K(x,y) \sum_i \frac{1}{\lambda_i} a_i \psi_i(x)$
- Set $K(x,y) = \frac{\sin(\pi(y-x))}{\sin\left(\frac{\pi(y-x)}{N}\right)} \to \lambda_i = 1$. K is the Dirichlet kernel.
  $f(y) = \sum_x \frac{\sin(\pi(y-x))}{\sin\left(\frac{\pi(y-x)}{N}\right)} \sum_i a_i \psi_i(x) = \sum_x \frac{\sin(\pi(y-x))}{\sin\left(\frac{\pi(y-x)}{N}\right)} f(x)$

Pattern recognition as Function extension problem

Definition

Let f be the classification function: $f(x) = \begin{cases} x \in C & 1 \\ x \notin C & -1 \end{cases}$

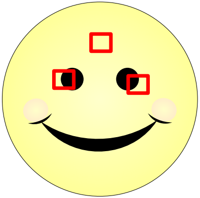
Definition

Given a kernel k with and its eigensystem $\{\psi_l, \lambda_l\}$ and a scalar function f

1. Compute the inner products $a_i = \langle \psi_l, f \rangle$
2. Extend the eigenfunction $\widetilde{\psi}_l(y) = \frac{1}{\lambda_l} \sum_{z \in X} k(y,z)\psi_l(z), y \notin X$
3. Extend f: $\widetilde{f} = \sum_l \langle \widetilde{\psi}_l, f \rangle_X \widetilde{\psi}_l, \forall y$

Example: eye detection I

Features extraction: SIFT [Lowe2003]



Learning

- We consider each SIFT descriptor as a sample in R^{128}.
- We collect a *learning set* of $\{P_i\}_1^N$ patches $\{f_i\}_1^N \in \{-1,1\}$.
- Embed $\{P_i\}_1^N$ using a density invariant embedding $\{\psi_l\}$.

Example: eye detection II

Recognition

- Given an input face, we extract the patches $\left\{\widehat{P}_k\right\}_1^{\widehat{N}}$.
- $\{\psi_l\}$ and $\{f_i\}_1^N$ are extended to $\left\{\widehat{P}_k\right\}_1^{\widehat{N}}$.
- $\{f_k\}_1^{\widehat{N}}$ is extended to $\left\{\widehat{P}_i\right\}_1^{\widehat{N}}$.
- The classification is given by $\begin{cases} f_k > 0 & \widehat{P}_k \in C \\ f_k < 0 & \widehat{P}_k \notin C \end{cases}$

Conclusions

1. Spectral embeddings give rise to class of adaptive bases
2. Functions can be extended over manifolds using such bases
3. Other applications (already implemented):
    - Image colorization
    - Sensor networks localization