On the Role of Eigendecomposition in Kernel Embedding

Jongha Jon Ryu, Jiun-Ting Huang, and Young-Han Kim

University of California, San Diego

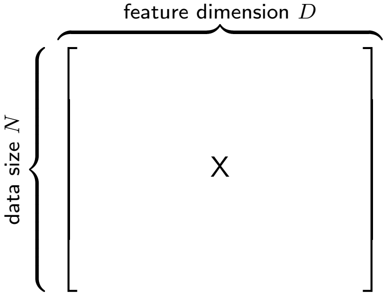
2021 IEEE International Symposium on Information Theory

Background

- We live in the era of **big** and **high-dimensional** data

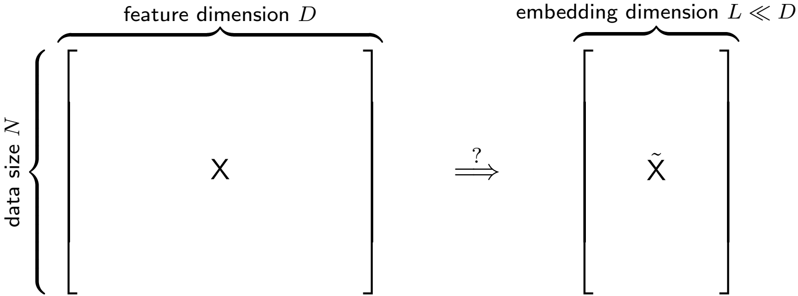
Background

- We live in the era of **big** and **high-dimensional** data



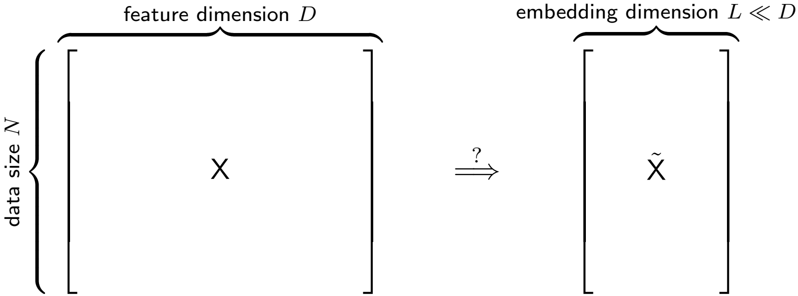
Background

- We live in the era of **big** and **high-dimensional** data
- **Desideratum**: A *fast* low-dimensional embedding method



Background

- We live in the era of **big** and **high-dimensional** data
- **Desideratum**: A *fast* low-dimensional embedding method



- **Applications**: Clustering, dimensionality reduction, visualization, ...

Motivation

- Kernel-based methods: EVD of $N \times N$ *kernel matrix*

Motivation

- Kernel-based methods: EVD of $N \times N$ *kernel matrix*
- e.g., Kernel PCA [5] and Laplacian eigenmaps [4, 1]

Motivation

- Kernel-based methods: EVD of $N \times N$ *kernel matrix*
- e.g., Kernel PCA [5] and Laplacian eigenmaps [4, 1]
- $(+)$ Simple, elegant, decent performance

Motivation

- Kernel-based methods: EVD of $N \times N$ *kernel matrix*
- e.g., Kernel PCA [5] and Laplacian eigenmaps [4, 1]
- $(+)$ Simple, elegant, decent performance
- $(-)$ Not easily scalable $O(N^3)$

Motivation

- Kernel-based methods: EVD of $N \times N$ *kernel matrix*
- e.g., Kernel PCA [5] and Laplacian eigenmaps [4, 1]
- $(+)$ Simple, elegant, decent performance
- $(-)$ Not easily scalable $O(N^3)$
- This work: **Kernel embedding without EVD of a matrix**

Motivation

- Kernel-based methods: EVD of $N \times N$ *kernel matrix*
- e.g., Kernel PCA [5] and Laplacian eigenmaps [4, 1]
- $(+)$ Simple, elegant, decent performance
- $(-)$ Not easily scalable $O(N^3)$
- This work: **Kernel embedding without EVD of a matrix**
  - For a special class of kernels, we only need density estimates

Motivation

- Kernel-based methods: EVD of $N \times N$ *kernel matrix*
- e.g., Kernel PCA [5] and Laplacian eigenmaps [4, 1]
- $(+)$ Simple, elegant, decent performance
- $(-)$ Not easily scalable $O(N^3)$
- This work: **Kernel embedding without EVD of a matrix**
  - For a special class of kernels, we only need density estimates
  - Example: dot-product kernels over hypersphere

Outline

1. **Preliminary**: PCA, kernel PCA and Laplacian eigenmaps
2. **Algorithm**: EVD-free kernel embedding with density oracle
3. **Example**: Dot-product kernels over hypersphere

Problem Setting and Notation

- Random vector $\mathbf{X} \sim p$ over a closed $\mathcal{X} \subset \mathbb{R}^d$

Problem Setting and Notation

- Random vector $\mathbf{X} \sim p$ over a closed $\mathcal{X} \subset \mathbb{R}^d$
- Data points $\mathbf{x}_{1:N} := \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \sim$ i.i.d. p

Problem Setting and Notation

- Random vector $\mathbf{X} \sim p$ over a closed $\mathcal{X} \subset \mathbb{R}^d$
- Data points $\mathbf{x}_{1:N} := \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \sim$ i.i.d. p
- Given *density* μ,

$$L_\mu^2(\mathcal{X}) := \left\{ f\colon \mathcal{X} \to \mathbb{C} \,\middle|\, \int |f(\mathbf{x})|^2 \,\mathrm{d}\mu(\mathbf{x}) < \infty \right\}$$

with $\langle f, g \rangle_\mu := \int f(\mathbf{x}) \overline{g(\mathbf{x})} \,\mathrm{d}\mu(\mathbf{x})$

Problem Setting and Notation

- Random vector $\mathbf{X} \sim p$ over a closed $\mathcal{X} \subset \mathbb{R}^d$
- Data points $\mathbf{x}_{1:N} := \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \sim$ i.i.d. p
- Given *density* μ,

$$L^2_\mu(\mathcal{X}) := \Big\{ f \colon \mathcal{X} \to \mathbb{C} \ \Big| \int |f(\mathbf{x})|^2 \, \mathrm{d}\mu(\mathbf{x}) < \infty \Big\}$$

with $\langle f, g \rangle_\mu := \int f(\mathbf{x}) \overline{g(\mathbf{x})} \, \mathrm{d}\mu(\mathbf{x})$

- A symmetric and compact kernel function $k \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

Problem Setting and Notation

- Random vector $\mathbf{X} \sim p$ over a closed $\mathcal{X} \subset \mathbb{R}^d$
- Data points $\mathbf{x}_{1:N} := \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \sim$ i.i.d. p
- Given *density* μ,

$$L^2_\mu(\mathcal{X}) := \left\{ f \colon \mathcal{X} \to \mathbb{C} \,\middle|\, \int |f(\mathbf{x})|^2 \,\mathrm{d}\mu(\mathbf{x}) < \infty \right\}$$

with $\langle f, g \rangle_\mu := \int f(\mathbf{x})\overline{g(\mathbf{x})}\,\mathrm{d}\mu(\mathbf{x})$

- A symmetric and compact kernel function $k \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
- The *associated Hilbert–Schmidt integral operator*

$$\mathbf{K} \colon L^2_\mu(\mathcal{X}) \to L^2_\mu(\mathcal{X}), \qquad (\mathbf{K}f)(\mathbf{x}) := \int_{\mathcal{X}} k(\mathbf{x}, \mathbf{t}) f(\mathbf{t}) \,\mathrm{d}\mu(\mathbf{t})$$

Principal Component Analysis

- PCA finds orthogonal directions $\mathbf{u}_1^\star, \ldots, \mathbf{u}_L^\star$ that capture variance of $\mathbf{X}$ as much as possible

- PCA embedding: $\psi_{\mathsf{PCA}}(\mathbf{x}) := [(\mathbf{u}_1^\star)^T \mathbf{x}, \ldots, (\mathbf{u}_L^\star)^T \mathbf{x}]^T$

Principal Component Analysis

- PCA finds orthogonal directions $\mathbf{u}_1^\star, \ldots, \mathbf{u}_L^\star$ that capture variance of $\mathbf{X}$ as much as possible
- PCA embedding: $\psi_{\mathsf{PCA}}(\mathbf{x}) := [(\mathbf{u}_1^\star)^T\mathbf{x}, \ldots, (\mathbf{u}_L^\star)^T\mathbf{x}]^T$
- Assume $\mathsf{E}\,\mathbf{X} = 0$

Principal Component Analysis

- PCA finds orthogonal directions $\mathbf{u}_1^\star, \ldots, \mathbf{u}_L^\star$ that capture variance of $\mathbf{X}$ as much as possible
- PCA embedding: $\psi_{\mathsf{PCA}}(\mathbf{x}) := [(\mathbf{u}_1^\star)^T\mathbf{x}, \ldots, (\mathbf{u}_L^\star)^T\mathbf{x}]^T$
- Assume $\mathsf{E}\,\mathbf{X} = 0$

Feature space (popul.)

$$\underset{\mathbf{u}_\ell \in \mathbb{R}^d}{\text{maximize}} \quad \sum_{\ell=1}^{L} \mathrm{Var}(\mathbf{u}_\ell^T \mathbf{X})$$
$$\text{subject to} \quad \mathbf{u}_\ell^T \mathbf{u}_{\ell'} = \delta_{\ell\ell'}$$

Principal Component Analysis

- PCA finds orthogonal directions $\mathbf{u}_1^\star, \ldots, \mathbf{u}_L^\star$ that capture variance of $\mathbf{X}$ as much as possible
- PCA embedding: $\psi_{\mathsf{PCA}}(\mathbf{x}) := [(\mathbf{u}_1^\star)^T\mathbf{x}, \ldots, (\mathbf{u}_L^\star)^T\mathbf{x}]^T$
- Assume $\mathsf{E}\,\mathbf{X} = 0$
- $\mathbf{C} := \mathrm{Cov}(\mathbf{X}, \mathbf{X}) = \mathsf{E}[\mathbf{X}\mathbf{X}^T]$

Feature space (popul.)

$$\begin{aligned} \underset{\mathbf{u}_\ell \in \mathbb{R}^d}{\text{maximize}} \quad & \sum_{\ell=1}^{L} \mathbf{u}_\ell^T \mathbf{C} \mathbf{u}_\ell \\ \text{subject to} \quad & \mathbf{u}_\ell^T \mathbf{u}_{\ell'} = \delta_{\ell\ell'} \end{aligned}$$

Principal Component Analysis

- PCA finds orthogonal directions $\mathbf{u}_1^\star, \ldots, \mathbf{u}_L^\star$ that capture variance of $\mathbf{X}$ as much as possible
- PCA embedding: $\psi_{\mathsf{PCA}}(\mathbf{x}) := [(\mathbf{u}_1^\star)^T\mathbf{x}, \ldots, (\mathbf{u}_L^\star)^T\mathbf{x}]^T$
- Assume $\mathsf{E}\,\mathbf{X} = 0$
- $\mathbf{C} := \mathrm{Cov}(\mathbf{X}, \mathbf{X}) = \mathsf{E}[\mathbf{X}\mathbf{X}^T]$
- $\hat{\mathbf{C}} := \frac{1}{N}\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i^T$ with samples

Feature space (sample)

$$\underset{\mathbf{u}_\ell \in \mathbb{R}^d}{\text{maximize}} \quad \sum_{\ell=1}^{L} \mathbf{u}_\ell^T \hat{\mathbf{C}} \mathbf{u}_\ell$$

$$\text{subject to} \quad \mathbf{u}_\ell^T \mathbf{u}_{\ell'} = \delta_{\ell\ell'}$$

Kernel PCA

- A feature map $|\phi(\cdot)\rangle \colon \mathcal{X} \to \mathcal{F}$, $\mathcal{F}$ a vector space with $\langle\cdot|\cdot\rangle$

Kernel PCA

- A feature map $|\phi(\cdot)\rangle \colon \mathcal{X} \to \mathcal{F}$, $\mathcal{F}$ a vector space with $\langle\cdot|\cdot\rangle$
- Kernel PCA $\equiv$ PCA on $|\phi(\mathbf{X})\rangle$

Kernel PCA

- A feature map $|\phi(\cdot)\rangle \colon \mathcal{X} \to \mathcal{F}$, $\mathcal{F}$ a vector space with $\langle\cdot|\cdot\rangle$
- Kernel PCA $\equiv$ PCA on $|\phi(\mathbf{X})\rangle$

Kernel PCA

- A feature map $|\phi(\cdot)\rangle\colon \mathcal{X} \to \mathcal{F}$, $\mathcal{F}$ a vector space with $\langle\cdot|\cdot\rangle$
- Kernel PCA $\equiv$ PCA on $|\phi(\mathbf{X})\rangle$

Feature space (popul.)

$$\begin{aligned} \underset{|u_\ell\rangle \in \mathcal{F}}{\text{maximize}} \quad & \sum_{\ell=1}^{L} \langle u_\ell | \mathbf{C}_\phi | u_\ell \rangle \\ \text{subject to} \quad & \langle u_\ell | u_{\ell'} \rangle = \delta_{\ell\ell'} \end{aligned}$$

Kernel PCA

- A feature map $|\phi(\cdot)\rangle\colon \mathcal{X} \to \mathcal{F}$, $\mathcal{F}$ a vector space with $\langle\cdot|\cdot\rangle$
- Kernel PCA $\equiv$ PCA on $|\phi(\mathbf{X})\rangle$ + kernel trick $k(\mathbf{x},\mathbf{t}) := \langle\phi(\mathbf{x})|\phi(\mathbf{t})\rangle$

Feature space (popul.)

$$\underset{|u_\ell\rangle\in\mathcal{F}}{\text{maximize}} \quad \sum_{\ell=1}^{L} \langle u_\ell|\mathbf{C}_\phi|u_\ell\rangle$$
$$\text{subject to} \quad \langle u_\ell|u_{\ell'}\rangle = \delta_{\ell\ell'}$$

Function space (popul.)

$$\underset{f_\ell\in L_p^2(\mathcal{X})}{\text{maximize}} \quad \sum_{\ell=1}^{L} \langle f_\ell, \mathbf{K}f_\ell\rangle_p$$
$$\text{subject to} \quad \langle f_\ell, f_{\ell'}\rangle_p = \delta_{\ell\ell'}$$

Kernel PCA

- A feature map $|\phi(\cdot)\rangle \colon \mathcal{X} \to \mathcal{F}$, $\mathcal{F}$ a vector space with $\langle \cdot | \cdot \rangle$
- Kernel PCA $\equiv$ PCA on $|\phi(\mathbf{X})\rangle$ + kernel trick $k(\mathbf{x}, \mathbf{t}) := \langle \phi(\mathbf{x}) | \phi(\mathbf{t}) \rangle$

Feature space (popul.)

$$\underset{|u_\ell\rangle \in \mathcal{F}}{\text{maximize}} \quad \sum_{\ell=1}^{L} \langle u_\ell | \mathbf{C}_\phi | u_\ell \rangle$$
$$\text{subject to} \quad \langle u_\ell | u_{\ell'} \rangle = \delta_{\ell\ell'}$$

Function space (popul.)

$$\underset{f_\ell \in L_p^2(\mathcal{X})}{\text{maximize}} \quad \sum_{\ell=1}^{L} \langle f_\ell, \mathbf{K} f_\ell \rangle_p$$
$$\text{subject to} \quad \langle f_\ell, f_{\ell'} \rangle_p = \delta_{\ell\ell'}$$

- (Population solution) top-L spectrum of $\mathbf{K}$ in $L_p^2(\mathcal{X})$
- (Population embedding)

$$\psi_{\mathsf{KPCA}}(\mathbf{x}) := [\sqrt{\lambda_1} f_1^\star(\mathbf{x}), \ldots, \sqrt{\lambda_L} f_L^\star(\mathbf{x})]^T$$

Kernel PCA

- A feature map $|\phi(\cdot)\rangle\colon \mathcal{X} \to \mathcal{F}$, $\mathcal{F}$ a vector space with $\langle\cdot|\cdot\rangle$
- Kernel PCA $\equiv$ PCA on $|\phi(\mathbf{X})\rangle$ + kernel trick $(\mathsf{K})_{mn} := k(\mathbf{x}_m, \mathbf{x}_n)$

Feature space (sample)

$$\underset{|u_\ell\rangle \in \mathcal{F}}{\text{maximize}} \quad \sum_{\ell=1}^{L} \langle u_\ell | \hat{\mathbf{C}}_\phi | u_\ell \rangle$$
$$\text{subject to} \quad \langle u_\ell | u_{\ell'} \rangle = \delta_{\ell\ell'}$$

Function space (sample)

$$\underset{\mathbf{f}_\ell \in \mathbb{R}^N}{\text{maximize}} \quad \sum_{\ell=1}^{L} \frac{\mathbf{f}_\ell^T}{\sqrt{N}} \frac{\mathsf{K}}{N} \frac{\mathbf{f}_\ell}{\sqrt{N}}$$
$$\text{subject to} \quad \frac{\mathbf{f}_\ell^T}{\sqrt{N}} \frac{\mathbf{f}_{\ell'}}{\sqrt{N}} = \delta_{\ell\ell'}$$

- (Sample solution) top-L spectrum of $\mathsf{K} \in \mathbb{R}^{N\times N}$ with $\mathbf{x}_{1:N} \sim$ i.i.d. p
- (Sample embedding)

$$\hat{\psi}_{\mathsf{KPCA}}(\mathbf{x}) = [\sqrt{\lambda_1}(\mathbf{f}_1^\star)_n, \ldots, \sqrt{\lambda_L}(\mathbf{f}_L^\star)_n]^T \quad \text{for } \mathbf{x} = \mathbf{x}_n$$

Laplacian Eigenmaps (a.k.a. Spectral Embedding)

- Given a base kernel function k, define

$$p_k(\mathbf{x}) := \int k(\mathbf{x},\mathbf{t})p(\mathbf{t})\,\mathrm{d}\mathbf{t} \quad \text{and} \quad \overline{k}_p(\mathbf{x},\mathbf{t}) := \frac{k(\mathbf{x},\mathbf{t})}{\sqrt{p_k(\mathbf{x})p_k(\mathbf{t})}}$$

Laplacian Eigenmaps (a.k.a. Spectral Embedding)

- Given a base kernel function k, define

$$p_k(\mathbf{x}) := \int k(\mathbf{x},\mathbf{t})p(\mathbf{t})\,\mathrm{d}\mathbf{t} \quad \text{and} \quad \overline{k}_p(\mathbf{x},\mathbf{t}) := \frac{k(\mathbf{x},\mathbf{t})}{\sqrt{p_k(\mathbf{x})p_k(\mathbf{t})}}$$

- LE $\equiv$ KPCA with the kernel $\overline{k}_p$ with embedding

$$\psi_{\mathsf{LE}}(\mathbf{x}) := [f_1^\star(\mathbf{x}), \ldots, f_L^\star(\mathbf{x})]^T$$

A New Density-regularized Kernel

- So far, reviewed KPCA and introduced LE as KPCA with $\overline{k}_p$

A New Density-regularized Kernel

- So far, reviewed KPCA and introduced LE as KPCA with $\overline{k}_p$
- Define a new kernel function

$$k_p(\mathbf{x}, \mathbf{t}) := \frac{k(\mathbf{x}, \mathbf{t})}{\sqrt{p(\mathbf{x})p(\mathbf{t})}}$$

and propose KPCA with k_p and embedding

$$\psi_{\mathsf{KE}}(\mathbf{x}) := [f_1^\star(\mathbf{x}), \ldots, f_L^\star(\mathbf{x})]^T$$

A New Density-regularized Kernel

- So far, reviewed KPCA and introduced LE as KPCA with $\overline{k}_p$
- Define a new kernel function

$$k_p(\mathbf{x}, \mathbf{t}) := \frac{k(\mathbf{x}, \mathbf{t})}{\sqrt{p(\mathbf{x})p(\mathbf{t})}}$$

and propose KPCA with k_p and embedding

$$\psi_{\mathsf{KE}}(\mathbf{x}) := [f_1^\star(\mathbf{x}), \ldots, f_L^\star(\mathbf{x})]^T$$

Function space (popul.)

$$\begin{aligned} \underset{f_\ell \in L_p^2(\mathcal{X})}{\text{maximize}} \quad & \sum_{\ell=1}^{L} \langle f_\ell, \mathbf{K}_p f_\ell \rangle_p \\ \text{subject to} \quad & \langle f_\ell, f_{\ell'} \rangle_p = \delta_{\ell\ell'} \end{aligned}$$

A New Density-regularized Kernel

- So far, reviewed KPCA and introduced LE as KPCA with $\overline{k}_p$
- Define a new kernel function

$$k_p(\mathbf{x}, \mathbf{t}) := \frac{k(\mathbf{x}, \mathbf{t})}{\sqrt{p(\mathbf{x})p(\mathbf{t})}}$$

and propose KPCA with k_p and embedding

$$\psi_{\mathsf{KE}}(\mathbf{x}) := [f_1^\star(\mathbf{x}), \ldots, f_L^\star(\mathbf{x})]^T$$

Function space (popul.)

$$\begin{aligned} &\underset{f_\ell \in L_p^2(\mathcal{X})}{\text{maximize}} \quad \sum_{\ell=1}^{L} \langle f_\ell, \mathbf{K}_p f_\ell \rangle_p \\ &\text{subject to} \quad \langle f_\ell, f_{\ell'} \rangle_p = \delta_{\ell\ell'} \end{aligned}$$

A New Density-regularized Kernel

- So far, reviewed KPCA and introduced LE as KPCA with $\overline{k}_p$
- Define a new kernel function

$$k_p(\mathbf{x}, \mathbf{t}) := \frac{k(\mathbf{x}, \mathbf{t})}{\sqrt{p(\mathbf{x})p(\mathbf{t})}}$$

and propose KPCA with k_p and embedding

$$\psi_{\mathsf{KE}}(\mathbf{x}) := [f_1^\star(\mathbf{x}), \ldots, f_L^\star(\mathbf{x})]^T$$

Function space (popul.)

$$\underset{f_\ell \in L_p^2(\mathcal{X})}{\text{maximize}} \quad \sum_{\ell=1}^{L} \langle f_\ell, \mathbf{K}_p f_\ell \rangle_p$$

$$\text{subject to} \quad \langle f_\ell, f_{\ell'} \rangle_p = \delta_{\ell\ell'}$$

Function space (sample)

$$\underset{\mathbf{f}_\ell \in \mathbb{R}^N}{\text{maximize}} \quad \sum_{\ell=1}^{L} \frac{\mathbf{f}_\ell^T}{\sqrt{N}} \frac{\mathsf{K}_p}{N} \frac{\mathbf{f}_\ell}{\sqrt{N}}$$

$$\text{subject to} \quad \frac{\mathbf{f}_\ell^T}{\sqrt{N}} \frac{\mathbf{f}_{\ell'}}{\sqrt{N}} = \delta_{\ell\ell'}$$

- **Claim**: We can directly solve the population problem for some cases!

A New Density-regularized Kernel

- So far, reviewed KPCA and introduced LE as KPCA with $\overline{k}_p$
- Define a new kernel function

$$k_p(\mathbf{x}, \mathbf{t}) := \frac{k(\mathbf{x}, \mathbf{t})}{\sqrt{p(\mathbf{x})p(\mathbf{t})}}$$

and propose KPCA with k_p and embedding

$$\psi_{\mathsf{KE}}(\mathbf{x}) := [f_1^\star(\mathbf{x}), \ldots, f_L^\star(\mathbf{x})]^T$$

Function space (popul.)

$$\underset{f_\ell \in L_p^2(\mathcal{X})}{\text{maximize}} \quad \sum_{\ell=1}^{L} \langle f_\ell, \mathbf{K}_p f_\ell \rangle_p$$
$$\text{subject to} \quad \langle f_\ell, f_{\ell'} \rangle_p = \delta_{\ell\ell'}$$

Neutralized ver.

$$\underset{g_\ell \in L_w^2(\mathcal{X})}{\text{maximize}} \quad \sum_{\ell=1}^{L} \langle g_\ell, \mathbf{K}_w g_\ell \rangle_w$$
$$\text{subject to} \quad \langle g_\ell, g_{\ell'} \rangle_w = \delta_{\ell\ell'}.$$

- **Claim**: We can directly solve the population problem for some cases!

EVD-free Kernel Embedding

Algorithm 1 EVD-free Kernel Embedding

Input base kernel k, weighting function w, sample $\{\mathbf{x}_n\}_{n=1}^N$, target dim. $L \in \mathbb{N}$, density estimator $\hat{p}(\cdot)$

Given The top-L orthonormal eigenfunctions $g_1^\star, \ldots, g_L^\star$ of the integral operator $\mathbf{K}_w \colon L_w^2(\mathcal{X}) \to L_w^2(\mathcal{X})$

1: Given a query point $\mathbf{x} \in \mathcal{X}$, output its L-dimensional embedding as

$$\hat{\psi}_{\mathsf{KE}}(\mathbf{x}) := \sqrt{\frac{w(\mathbf{x})}{\hat{p}(\mathbf{x})}}[g_1^\star(\mathbf{x}), \ldots, g_L^\star(\mathbf{x})]^T$$

- Note: the final embedding is in the flavor of LE embedding
- Is there really such a nice kernel? YES!

Dot-product Kernels over Hypersphere

- **A special class of kernels**: *Dot-product kernels* [6]

$$k_w(\mathbf{x}, \mathbf{t}) = f(\mathbf{x}^T \mathbf{t})$$

for some function $f\colon \mathbb{R} \to \mathbb{R}$

Dot-product Kernels over Hypersphere

- **A special class of kernels**: *Dot-product kernels* [6]

$$k_w(\mathbf{x}, \mathbf{t}) = f(\mathbf{x}^T \mathbf{t})$$

for some function $f \colon \mathbb{R} \to \mathbb{R}$

- **A special domain**: the unit hypersphere

$$\mathbb{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d \colon \|\mathbf{x}\|_2 = 1\} \subset \mathbb{R}^d$$

Dot-product Kernels over Hypersphere

- **A special class of kernels**: *Dot-product kernels* [6]

$$k_w(\mathbf{x}, \mathbf{t}) = f(\mathbf{x}^T \mathbf{t})$$

  for some function $f\colon \mathbb{R} \to \mathbb{R}$

- **A special domain**: the unit hypersphere

$$\mathbb{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d \colon \|\mathbf{x}\|_2 = 1\} \subset \mathbb{R}^d$$

- **Examples**: Gaussian kernels $f(u) = e^{-(1+u)/\sigma^2}$ $(\sigma > 0)$,
  arccosine kernel $f(u) = 1 - (2/\pi)\cos^{-1}(u)$, ...

Dot-product Kernels over Hypersphere

- **A special class of kernels**: *Dot-product kernels* [6]

$$k_w(\mathbf{x}, \mathbf{t}) = f(\mathbf{x}^T \mathbf{t})$$

for some function $f\colon \mathbb{R} \to \mathbb{R}$

- **A special domain**: the unit hypersphere

$$\mathbb{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d \colon \|\mathbf{x}\|_2 = 1\} \subset \mathbb{R}^d$$

- **Examples**: Gaussian kernels $f(u) = e^{-(1+u)/\sigma^2}$ $(\sigma > 0)$,
  arccosine kernel $f(u) = 1 - (2/\pi)\cos^{-1}(u)$, ...

- **Key property**: with uniform weighting function w,
  spherical harmonics fully characterize the eigensystem of $\mathbf{K}_w$!

Other Examples Beyond Hypersphere

- Multiplicative dot-product kernels over a torus
- Dot-product kernels over a ball
- Gaussian kernels with Gaussian weighting

Experiment: Patch-based Image Segmentation

- Orders-of-magnitude faster ($\sim$2s) than other methods ($\sim$100s)

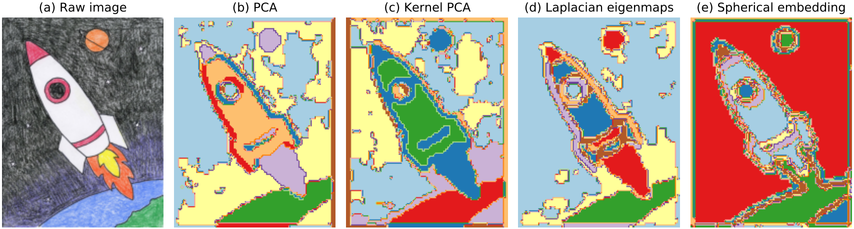


Figure: An illustrative example with image segmentation

Concluding Remarks

- Equivalence between KPCA and LE have been long known [4, 3, 2]

Concluding Remarks

- Equivalence between KPCA and LE have been long known [4, 3, 2]
- An unorthodox view on kernel-based embeddings based on the population optimization perspective

Concluding Remarks

- Equivalence between KPCA and LE have been long known [4, 3, 2]
- An unorthodox view on kernel-based embeddings based on the population optimization perspective
- The proposed EVD-free kernel embedding algorithm can be an extremely low-cost kernel-based embedding

Concluding Remarks

- Equivalence between KPCA and LE have been long known [4, 3, 2]
- An unorthodox view on kernel-based embeddings based on the population optimization perspective
- The proposed EVD-free kernel embedding algorithm can be an extremely low-cost kernel-based embedding
- Many interesting open questions

Concluding Remarks

- Equivalence between KPCA and LE have been long known [4, 3, 2]
- An unorthodox view on kernel-based embeddings based on the population optimization perspective
- The proposed EVD-free kernel embedding algorithm can be an extremely low-cost kernel-based embedding
- Many interesting open questions
  - Any theoretical justification as for LE?

Concluding Remarks

- Equivalence between KPCA and LE have been long known [4, 3, 2]
- An unorthodox view on kernel-based embeddings based on the population optimization perspective
- The proposed EVD-free kernel embedding algorithm can be an extremely low-cost kernel-based embedding
- Many interesting open questions
  - Any theoretical justification as for LE?
  - Universally comparable practical performance?

References I

Mikhail Belkin and Partha Niyogi.
Laplacian eigenmaps for dimensionality reduction and data representation.
Neural Comput., 15(6):1373–1396, 2003.

Yoshua Bengio, Olivier Delalleau, Nicolas Le Roux, Jean-François Paiement, Pascal Vincent, and Marie Ouimet.
Learning eigenfunctions links spectral embedding and kernel pca.
Neural Comput., 16(10):2197–2219, 2004.

Jihun Ham, Daniel D Lee, Sebastian Mika, and Bernhard Schölkopf.
A kernel view of the dimensionality reduction of manifolds.
In *Proc. Int. Conf. Mach. Learn.*, page 47, 2004.

Andrew Ng, Michael Jordan, and Yair Weiss.
On spectral clustering: Analysis and an algorithm.
Adv. Neural Inf. Proc. Syst., 14:849–856, 2001.

References II

Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller.
Nonlinear component analysis as a kernel eigenvalue problem.
Neural Comput., 10(5):1299–1319, 1998.

Alex J Smola, Zoltan L Ovari, and Robert C Williamson.
Regularization with dot-product kernels.
Adv. Neural Inf. Proc. Syst., pages 308–314, 2001.