Kernel PCA and De-Noising in Feature Spaces

Presented by

Jökull Jóhannsson
Kristófer Hannesson
Xinyi Lin
Sudhanshu Mittal
Samuel Murray

Kernel PCA and De-Noising in Feature Spaces

- Linear/Kernel PCA
- The method
- Experiment results
- Discussion

Linear vs Kernel PCA

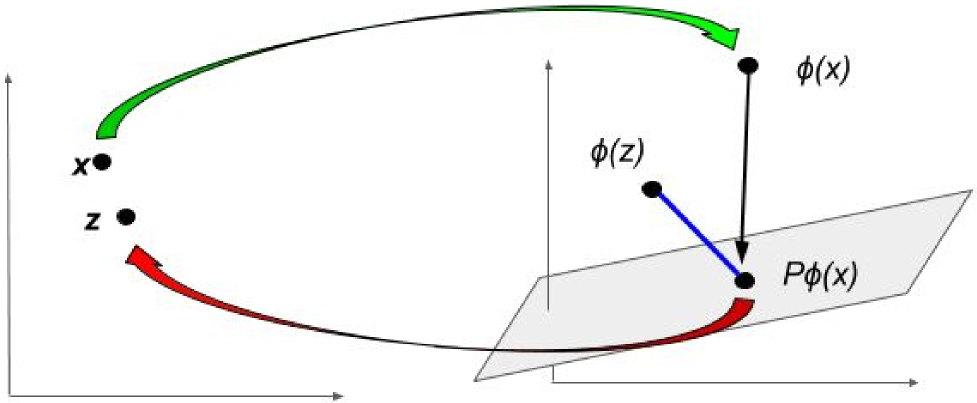
Applications

- compression
- reconstruction
- de-noising

What we explored

- Nonlinear de-noising
- Connection between feature space expansions and meaningful patterns in input space.

Linear vs Kernel PCA



Kernel PCA

- Map to feature space using function Φ
- Do PCA in the feature space - calculate projection P
- Results may not have pre-images in the input space
  - Solution: Find approximate pre-image (z)

Methodology

$$N\lambda\alpha = K\alpha,$$

PCA

K = C, N=1

$$k(x, y) = \exp(-||x-y||^2/c)$$

$$\beta_k := V^k \cdot \Phi(x)$$

$$P_n\Phi(x) = \sum_{k=1}^{n} \beta_k V^k$$

Input Space

k - Gaussian kernel function

z - pre-image of vector x in input space

ϕ - mapping

Feature Space

α - eigenvectors

V - normalized eigenvectors

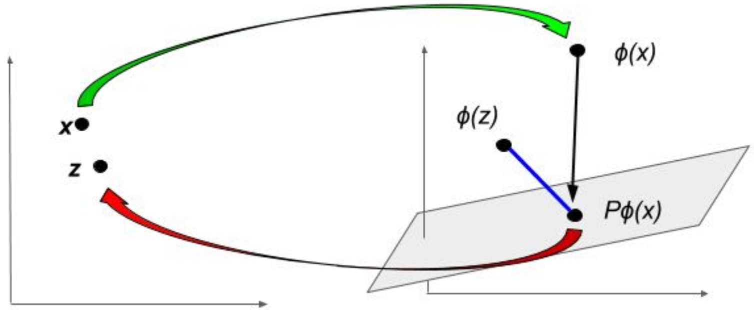
β - projection on principal components

γ - weight of each training point

n - number of top components

Pre-images for Gaussian Kernels

$$\gamma_i = \sum_{k=1}^{n} \beta_k \alpha_i^k .$$



Minimize : $\rho(z) = ||\Phi(z) - P_n \Phi(x)||^2$

$$\nabla_z \rho(z) = 0$$

Using Fixed-point Iterative method, we calculate:

$$z_{t+1} = \frac{\sum_{i=1}^{N} \gamma_i \exp(-||z_t - x_i||^2/c) x_i}{\sum_{i=1}^{N} \gamma_i \exp(-||z_t - x_i||^2/c)}$$

Experiments

- Toy examples
- Reconstruction and denoising of digits

Toy example: Denoising in $\mathbb{R}^2$

Linear PCA

- 1 or 2 components
- 2 components reproduces all noise!

Toy example: Denoising in $\mathbb{R}^2$

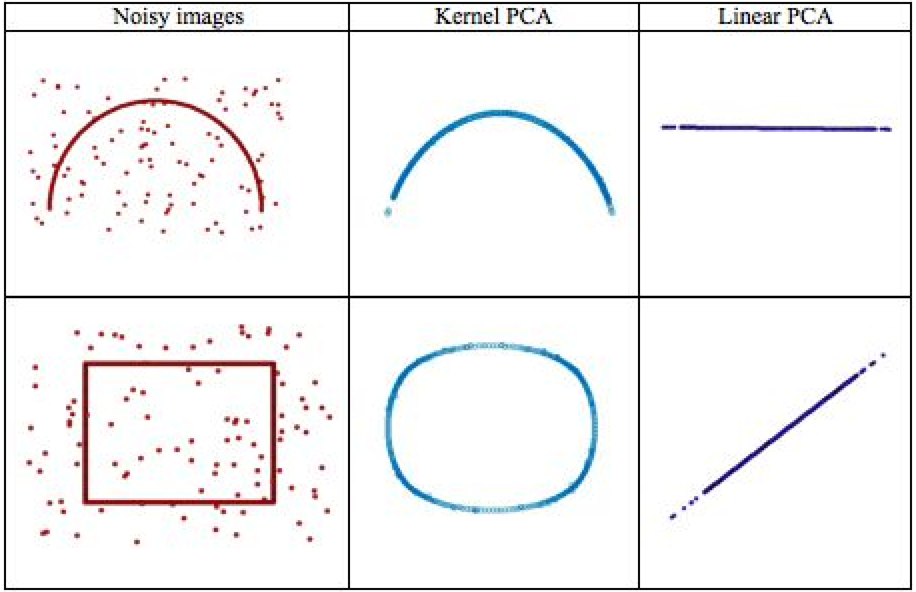
Linear PCA

- 1 or 2 features
- 2 features reproduces all noise!

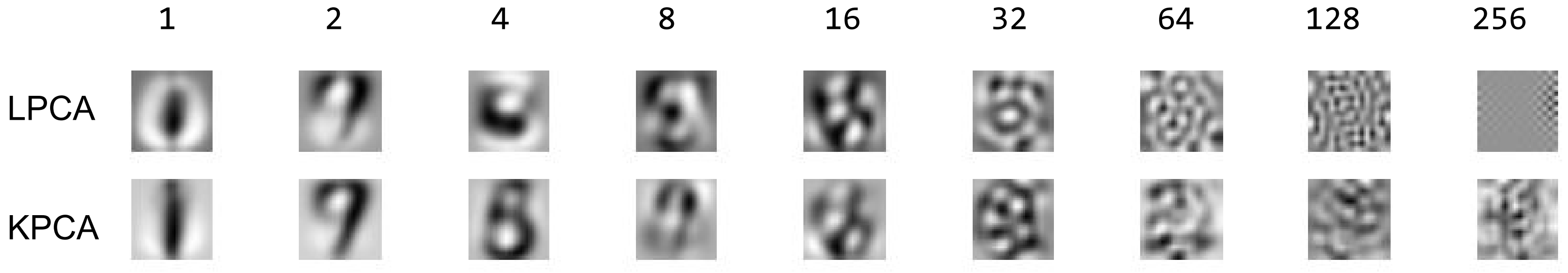
Kernel PCA

- As many components as training points!

Toy example: Denoising in $\mathbb{R}^2$



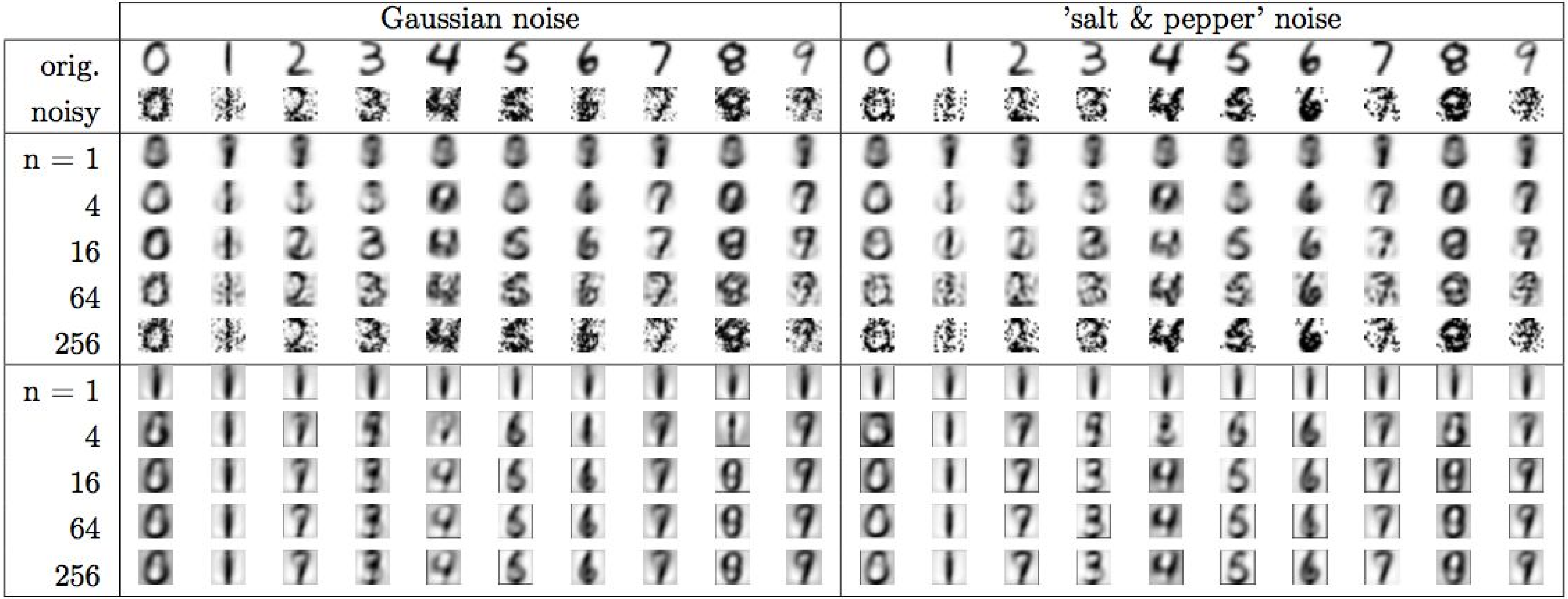
Eigenvectors for components $2^0,\ldots,2^8$



Reconstruction

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Denoising comparison



Discussion

Advantages

- Captures complex nonlinear structure in the data
- Allows projection of data to high dimensional space
- Can be used for application of reconstruction, denoising and compression.

Disadvantages

- Suffers from numerical instability or local minima problem.
- Dependent on initial guess.
- For the fixpoint method, the kernel function needs to be smooth.
- Computationally intensive compared to PCA

Summary

- Performs PCA on data in feature space
- Iterative process to get pre-image
- Can de-noise and reconstruct non-linear data

THANK YOU