Unsupervised Learning Methods
Problem Set III –
PCA and KPCA

Due: 16.05.2022

Guidelines

- Answer all questions (PDF + Jupyter notebook).
- You must type your solution manual (handwriting is not allowed).
- Submission in pairs (use the forum if needed).
- You **may** submit the entire solution in a single ipynb file (or in PDF + ipynb files).
- You **may** (and should) use the forum if you have any questions.
- Good luck!

1 Eigendecomposition

- Let $\boldsymbol{A} \in \mathbb{R}^{d \times d}$ be a square matrix.
- Let $\boldsymbol{u} \in \mathbb{R}^{d}$ be an eigenvector with eigenvalue λ, that is, $\boldsymbol{A}\boldsymbol{u} = \lambda \boldsymbol{u}$.
- Let $\boldsymbol{v} = \alpha \boldsymbol{u}$ where $\alpha \neq 0$.

1.1

Is $\boldsymbol{v}$ an eigenvector of $\boldsymbol{A}$?
If so, find its corresponding eigenvalue.
If not, give a counterexample.

Solution:
Type your solution here...

- Let $\boldsymbol{A} \in \mathbb{R}^{d \times d}$ be a square matrix.
- Let $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{R}^{d}$ where $\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle = 0$ be eigenvectors with the same eigenvalue λ, that is, $\boldsymbol{A}\boldsymbol{u}_1 = \lambda \boldsymbol{u}_1$ and $\boldsymbol{A}\boldsymbol{u}_2 = \lambda \boldsymbol{u}_2$.

1.2

Prove or disprove:
$\boldsymbol{u} = \boldsymbol{u}_1 + \boldsymbol{u}_2$ is an eigenvector of $\boldsymbol{A}$.

Solution:
Type your solution here...

- Let $\boldsymbol{A} \in \mathbb{R}^{d \times d}$ be a diagonalizable matrix, that is, $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}$ where $\boldsymbol{\Lambda}$ is a diagonal matrix.

1.3

Show that:

$$\operatorname{Tr}\left\{\boldsymbol{A}\right\} = \sum_{i=1}^{d} \lambda_i\left(\boldsymbol{A}\right)$$

where $\lambda_i\left(\boldsymbol{A}\right) = \boldsymbol{\Lambda}\left[i, i\right]$ is the ith eigenvalue of $\boldsymbol{A}$.

Solution:
Type your solution here...

- Let $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ be a diagonalizable matrix.
- Let $\{(\boldsymbol{u}_i, \lambda_i)\}_{i=1}^d$ be the set of eigen-pairs, that is, $\boldsymbol{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$.
- Let $\widetilde{\boldsymbol{A}} = \boldsymbol{R}\boldsymbol{A}\boldsymbol{R}^T$ where $\boldsymbol{R} \in \mathbb{R}^{d\times d}$ is orthogonal, that is $\boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{I}_d$.

1.4

Find $\left\{\left(\tilde{\boldsymbol{u}}_i, \tilde{\lambda}_i\right)\right\}_{i=1}^d$ the set of eigen-pairs of $\widetilde{\boldsymbol{A}}$ (such that: $\widetilde{\boldsymbol{A}}\tilde{\boldsymbol{u}}_i = \tilde{\lambda}_i \tilde{\boldsymbol{u}}_i$).

Solution:
Type your solution here...

Similarity

- Two (square) matrices $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ and $\boldsymbol{B} \in \mathbb{R}^{d\times d}$ are called similar, namely, $\boldsymbol{A} \sim \boldsymbol{B}$, if exists an (invertible) matrix $\boldsymbol{P} \in \mathbb{R}^{d\times d}$ such that:

$$\boldsymbol{B} = \boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}$$

1.5

Prove that if $\boldsymbol{A}$ is diagonalizable and $\boldsymbol{A} \sim \boldsymbol{B}$, then, $\boldsymbol{A}$ and $\boldsymbol{B}$ share the same set of eigenvalues, namely:

$$\boldsymbol{A} \sim \boldsymbol{B} \implies \{\lambda_i(\boldsymbol{A})\}_{i=1}^d = \{\lambda_i(\boldsymbol{B})\}_{i=1}^d$$

Solution:
Type your solution here...

SPD matrices
A symmetric matrix $\boldsymbol{S} = \boldsymbol{S}^T$ is an Symmetric Positive Definite (SPD), namely $\boldsymbol{S} \succ 0$ if either:

1. $\lambda_i(\boldsymbol{S}) > 0$ for all i.
2. $\boldsymbol{v}^T\boldsymbol{S}\boldsymbol{v} > 0$ for all $\boldsymbol{v} \neq \boldsymbol{0}$.

1.6

Prove that the two conditions are equivalent, that is:

$$\lambda_i(\boldsymbol{S}) > 0 \iff \boldsymbol{v}^T\boldsymbol{S}\boldsymbol{v} > 0 \quad \forall \boldsymbol{v} \neq \boldsymbol{0}$$

Solution:
Type your solution here...

- Let $\boldsymbol{S} \succ 0$ be an SPD matrix.

1.7

Prove or disprove:
$\boldsymbol{S}^{-1}$ is an SPD matrix.

Solution:
Type your solution here...

- Let $\boldsymbol{S} \succ 0$ be an SPD matrix.

1.8

Prove or disprove:
The SVD $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ is also an eigendecomposition.

Solution:
Type your solution here...

---

2 PCA

Full PCA

- Consider the data $\mathcal{X} = \left\{\boldsymbol{x}_i \in \mathbb{R}^D\right\}_{i=1}^N$ with mean $\boldsymbol{\mu}_x \in \mathbb{R}^D$ and covariance $\boldsymbol{\Sigma}_x \in \mathbb{R}^{D \times D}$.
- Let $\boldsymbol{\Sigma}_x = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ be the eigendecomposition of $\boldsymbol{\Sigma}_x$.
- Let $\boldsymbol{z}_i = \boldsymbol{U}^T(\boldsymbol{x}_i - \boldsymbol{\mu}_x)$.

2.1

Prove that:

1. The mean of $\mathcal{Z} = \{\boldsymbol{z}_i\}_{i=1}^N$ is zero, that is, $\boldsymbol{\mu}_z = \frac{1}{N}\sum_{i=1}^N \boldsymbol{z}_i = \boldsymbol{0}$.
2. The covariance of $\mathcal{Z}$ is diagonal, that is $\boldsymbol{\Sigma}_z$ is diagonal.
3. $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 = \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2$ for all i and j .

Solution:
Type your solution here...

- Consider the data $\{\boldsymbol{x}_i \in \mathbb{R}^{10}\}_{i=1}^{10}$ and its corresponding matrix $\boldsymbol{X} \in \mathbb{R}^{10\times N}$.
- Let $\boldsymbol{\Sigma}_x$ be the covariance matrix where all eigenvalues of $\boldsymbol{\Sigma}_x$ are unique.
- Let $\{\boldsymbol{z}_i \in \mathbb{R}^3\}_{i=1}^{N}$ be the low-dimensional representation obtained by applying PCA from $\mathbb{R}^{10}$ to $\mathbb{R}^3$.
- Let $\{\tilde{\boldsymbol{z}}_i \in \mathbb{R}^2\}_{i=1}^{N}$ be the low-dimensional representation obtained by applying PCA from $\mathbb{R}^{10}$ to $\mathbb{R}^2$.

2.2

Prove or disprove:

$$\tilde{\boldsymbol{z}}_i = \begin{bmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \end{bmatrix} \boldsymbol{z}_i$$

where $\alpha_1, \alpha_2 \in \{-1, 1\}$.

Solution:
Type your solution here...

Geometric PCA

- Let $\boldsymbol{U}_d \in \mathbb{R}^{D\times d}$ be a full rank matrix (with $d \leq D$).

2.3

Show that exists an invertible matrix $\boldsymbol{M} \in \mathbb{R}^{d\times d}$ such that $\boldsymbol{O} = \boldsymbol{U}_d\boldsymbol{M} \in \mathbb{R}^{D\times d}$ is semi-orthogonal, that is:

$$\boldsymbol{O}^T\boldsymbol{O} = \boldsymbol{I}_d$$

Hint: Use SVD.

Solution:
Type your solution here...

- Consider the data $\boldsymbol{X} \in \mathbb{R}^{D \times N}$ with zero mean $\boldsymbol{X}\boldsymbol{1}_N = \boldsymbol{0} \in \mathbb{R}^D$ and covariance $\boldsymbol{\Sigma}_x = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^T \in \mathbb{R}^{D \times D}$.
- Consider the following optimization problems:

1. Reconstruction error minimization:

$$\begin{cases} \arg\min_{\boldsymbol{U}_d \in \mathbb{R}^{D \times d}} \left\| \boldsymbol{X} - \boldsymbol{U}_d \boldsymbol{U}_d^T \boldsymbol{X} \right\|_F^2 \\ \text{s.t. } \boldsymbol{U}_d^T \boldsymbol{U}_d = \boldsymbol{I}_d \end{cases}$$

2. Variance maximization:

$$\begin{cases} \arg\max_{\boldsymbol{U}_d \in \mathbb{R}^{D \times d}} \operatorname{Tr}\left\{ \boldsymbol{U}_d^T \boldsymbol{\Sigma}_x \boldsymbol{U}_d \right\} \\ \text{s.t. } \boldsymbol{U}_d^T \boldsymbol{U}_d = \boldsymbol{I}_d \end{cases}$$

2.4

Prove that both problems have the same optimal solution $\boldsymbol{U}_d^\star$.

Solution:
Type your solution here...

PCA analysis

- Consider the data $\left\{ \boldsymbol{x}_i \in \mathbb{R}^D \right\}_{i=1}^N$ with mean $\boldsymbol{\mu}_x \in \mathbb{R}^D$ and covariance $\boldsymbol{\Sigma}_x \in \mathbb{R}^{D \times D}$.
- Let $\boldsymbol{U}_d \in \mathbb{R}^{D \times d}$ be a semi-orthogonal matrix, that is, $\boldsymbol{U}_d^T \boldsymbol{U}_d = \boldsymbol{I}_d$.
- Let $\boldsymbol{z}_i = \boldsymbol{U}_d^T \left(\boldsymbol{x}_i - \boldsymbol{\mu}_x \right) \in \mathbb{R}^d$.
- Let $\hat{\boldsymbol{x}}_i = \boldsymbol{U}_d \boldsymbol{z}_i + \boldsymbol{\mu}_x \in \mathbb{R}^D$.
- Let $\boldsymbol{\epsilon}_i = \boldsymbol{x}_i - \hat{\boldsymbol{x}}_i \in \mathbb{R}^D$.

2.5

Prove that:

$$\operatorname{Tr}\left\{ \boldsymbol{\Sigma}_x \right\} = \operatorname{Tr}\left\{ \boldsymbol{\Sigma}_z \right\} + \operatorname{Tr}\left\{ \boldsymbol{\Sigma}_\epsilon \right\}$$

where:

- $\boldsymbol{\Sigma}_z \in \mathbb{R}^{d \times d}$ is the covariance of $\left\{ \boldsymbol{z}_i \right\}_{i=1}^N$.
- $\boldsymbol{\Sigma}_\epsilon \in \mathbb{R}^{D \times D}$ is the covariance of $\left\{ \boldsymbol{\epsilon}_i \right\}_{i=1}^N$.

Solution:
Type your solution here...

- Let $\boldsymbol{U}_d \in \mathbb{R}^{D \times d}$ be the top d eigenvectors corresponding to the d largest eigenvalues of $\boldsymbol{\Sigma}_x$.

2.6

Show that:

$$\operatorname{Tr}\left\{\boldsymbol{\Sigma}_\epsilon\right\} = \sum_{i=d+1}^{D} \lambda_i\left(\boldsymbol{\Sigma}_x\right)$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_D$

Solution:
Type your solution here...

High-dimensional data PCA

- Consider the data $\boldsymbol{X} \in \mathbb{R}^{D \times N}$ where $D > N$.
- Let $\boldsymbol{\Sigma}_x \in \mathbb{R}^{D \times D}$ be the empirical covariance.

2.7

- Provide a (tight) upper bound on the number of non-zero eigenvalues of $\boldsymbol{\Sigma}_x$.
- Consequently, can you apply PCA to $\boldsymbol{X} \in \mathbb{R}^{D \times N}$ to obtain $\boldsymbol{Z} \in \mathbb{R}^{d \times N}$ with $d < D$ such that there is no loss of information?
  Explain your answer.

Solution:
Type your solution here...

Rank minimization

- Let $\boldsymbol{A} \in \mathbb{R}^{D \times N}$.
- Consider the following rank minimization problem:

$$\begin{cases} \min_{\boldsymbol{M} \in \mathbb{R}^{D \times N}} \|\boldsymbol{A} - \boldsymbol{M}\|_F^2 \\ \text{s.t. } \operatorname{rank}(\boldsymbol{M}) \leq d \end{cases}$$

2.8

- Solve the optimization problem.
- Write your final solution using the (truncated) matrices obtained by the SVD decomposition of $\boldsymbol{A}$, namely, $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$.

Hints:

1. Any matrix $\boldsymbol{M} \in \mathbb{R}^{D \times N}$ with $\operatorname{rank}(\boldsymbol{M}) = d$ can be written as $\boldsymbol{M} = \boldsymbol{B}\boldsymbol{C}$ where:
   (a) $\boldsymbol{B} \in \mathbb{R}^{D \times d}$
   (b) $\boldsymbol{C} \in \mathbb{R}^{d \times N}$

   use this result to formulate (and solve) an equivalent unconstrained problem.
2. There is a strong connection to PCA (no need to start from scratch).

Solution:
Type your solution here...

2.9 Implementation and applications

Solve this section in the attached notebook.

3 KPCA

Centering matrix

- Let $\boldsymbol{J} = \boldsymbol{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^T \in \mathbb{R}^{N \times N}$ be the centering matrix.

3.1

Prove that $\boldsymbol{J}$ is idempotent, that is, $\boldsymbol{J}^2 = \boldsymbol{J}$.
In words, applying $\boldsymbol{J}$ twice gives the same results as applying it once.

Solution:
Type your solution here...

Kernel matrix

- Let $\boldsymbol{X} \in \mathbb{R}^{D \times N}$ and let:

$$\boldsymbol{\Sigma}_x := \boldsymbol{X}\boldsymbol{X}^T$$
$$\boldsymbol{K}_x := \boldsymbol{X}^T\boldsymbol{X}$$

Let $(\boldsymbol{u}_i, \lambda_i)$ be an eigen-pair of $\boldsymbol{\Sigma}_x$ such that $\boldsymbol{\Sigma}_x \boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$ with $\lambda_i > 0$.

3.2

1. Show that λ_i is an eigenvalue of $\boldsymbol{K}_x$ as well.
2. Find its corresponding eigenvector such that $\boldsymbol{K}_x \boldsymbol{w}_i = \lambda_i \boldsymbol{w}_i$.

Solution:
Type your solution here...

Kernel functions

- Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and consider $\left\{\boldsymbol{x}_i \in \mathbb{R}^d\right\}_{i=1}^N$.

3.3

Show that if k can be written as an inner product, that is

$$k\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \left\langle \phi\left(\boldsymbol{x}_i\right), \phi\left(\boldsymbol{x}_j\right) \right\rangle$$

for some ϕ, then, the matrix defined by:

$$\boldsymbol{K}_x\left[i, j\right] = k\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)$$

is an SPSD matrix, namely, $\boldsymbol{K}_x \succeq 0$.

Solution:
Type your solution here...

- Let $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathbb{R}^d$ and consider:

$$k_1\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \left(1 + \boldsymbol{x}_i^T \boldsymbol{x}_j\right)^2$$

$$k_2\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \boldsymbol{x}_i^T \boldsymbol{A} \boldsymbol{x}_j$$

  where $\boldsymbol{A} \succ 0$.

3.4

Prove or disprove:

1. k_1 is a kernel function.
2. k_2 is a kernel function.

Solution:
Type your solution here...

- Consider $\left\{\boldsymbol{x}_i \in \mathbb{R}^d\right\}_{i=1}^N$, and consider the kernel:

$$k\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) := \left\langle \boldsymbol{\phi}_i, \boldsymbol{\phi}_j \right\rangle$$

  where $\boldsymbol{\phi}_i = \phi\left(\boldsymbol{x}_i\right)$ for all i for some ϕ.

- Let:

$$\tilde{k}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) := \left\langle \boldsymbol{\phi}_i - \boldsymbol{\mu}_\phi, \boldsymbol{\phi}_j - \boldsymbol{\mu}_\phi \right\rangle$$

  be the centered version, where:

$$\boldsymbol{\mu}_\phi = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\phi}_i$$

3.5

Show that $\tilde{k}$ can be written using only k, without using ϕ and $\boldsymbol{\mu}_\phi$ explicitly.

Solution:
Type your solution here...

- -

- Let $\boldsymbol{K}_x \in \mathbb{R}^{N \times N}$ be a kernel matrix, that is:

$$\boldsymbol{K}_x\left[i, j\right] = k\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)$$

  for some kernel function k.

- Let $\widetilde{\boldsymbol{K}}_x$ be the centered version, that is:

$$\widetilde{\boldsymbol{K}}_x = \boldsymbol{J} \boldsymbol{K}_x \boldsymbol{J}$$

  where $\boldsymbol{J} = \boldsymbol{I} - \frac{1}{N} \mathbf{1} \mathbf{1}^T$.

3.6

Prove or disprove:
$\widetilde{\boldsymbol{K}}_x$ is an SPD matrix (that is, all eigenvalues are strictly positive).

Solution:
Type your solution here...

- -

Out of sample extension

- Let $\boldsymbol{K}_x \in \mathbb{R}^{N\times N}$ be the kernel matrix obtained from the training set $\mathcal{X} = \left\{\boldsymbol{x}_i \in \mathbb{R}^D\right\}_{i=1}^N$.
- Let $\boldsymbol{Z} \in \mathbb{R}^{d\times N}$ be the low-dimensional representation obtained by applying KPCA, that is:

$$\boldsymbol{Z} = \boldsymbol{\Sigma}_d \boldsymbol{V}_d^T$$

  where $\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^T = \boldsymbol{J}\boldsymbol{K}_x\boldsymbol{J}$ is an eigendecomposition (see lecture notes).
- Let $\boldsymbol{X}^\star \in \mathbb{R}^{D\times N^\star}$ be a set of new unseen data-points.

3.7

Write an expression (in a matrix form) for $\boldsymbol{Z}^\star \in \mathbb{R}^{d\times N^\star}$, the KPCA out of sample extension applied to $\boldsymbol{X}^\star$.

Solution:
Type your solution here...

– –

- Let $\mathcal{X} = \{\boldsymbol{x}_i\}_{i=1}^N$ be the training set.
- Let $\mathcal{Z} = \{\boldsymbol{z}_i\}_{i=1}^N$ be the representation obtained by KPCA (training encoding).
- Consider a new point $\boldsymbol{x}^\star$ where $\boldsymbol{x}^\star = \boldsymbol{x}_k$ for some $k \leq N$.
- Let $\boldsymbol{z}^\star$ be the out of sample encoding applied to $\boldsymbol{x}^\star$.

3.8

Prove of disprove:

$$\boldsymbol{z}^\star = \boldsymbol{z}_k$$

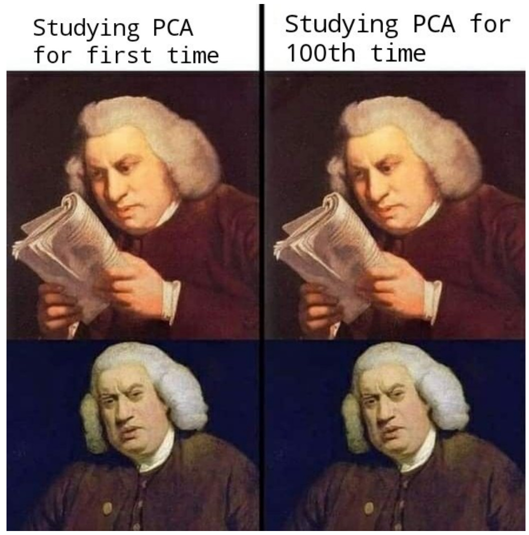
Solution:
Type your solution here...

– –

3.9 Implementation and applications

Solve this section in the attached notebook.

– –


Studying PCA for first time
Studying PCA for 100th time