ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Models, Algorithms and Data (MAD): Introduction to Computing

Spring semester 2021

Prof. Dr. Jens Honore Walther
Dr. Georgios Arampatzis
ETH Zentrum, CLT
CH-8092 Zürich

Set 11

Issued: 28.05.2021

In this exercise, you will learn how to perform dimensionality reduction using Principle Component Analysis (PCA). The accompanying notebook will allow you to apply this knowledge in a practical setting.

Question 1: Dimensionality Reduction with PCA

In the following, we want to see how we can use PCA to perform dimensionality reduction. You are given a dataset of measurements $\mathbf{x}_i \in \mathbb{R}^D$ for $i = 1, \ldots, N$. First, we want to transform the data such that is has zero mean (i.e. $\sum \mathbf{x}_i^{cent} = 0$ for $\mathbf{x}_i^{cent} = \mathbf{x}_i - \bar{x}$, where $\bar{x} = \frac{1}{N}\sum \mathbf{x}_i$), and then combine it in a matrix $X_{cent} = (\mathbf{x}_1^{cent}, \ldots, \mathbf{x}_n^{cent})^\top \in \mathbb{R}^{N \times D}$. The covariance matrix $\hat{C} \in \mathbb{R}^{D \times D}$ of this data is written as

$$\hat{C} = \frac{1}{N-1} X_{cent}^\top X_{cent} \tag{1}$$

In the lecture, you saw that finding the direction in the data which maximizes its variance is obtained by performing an eigenvalue decomposition

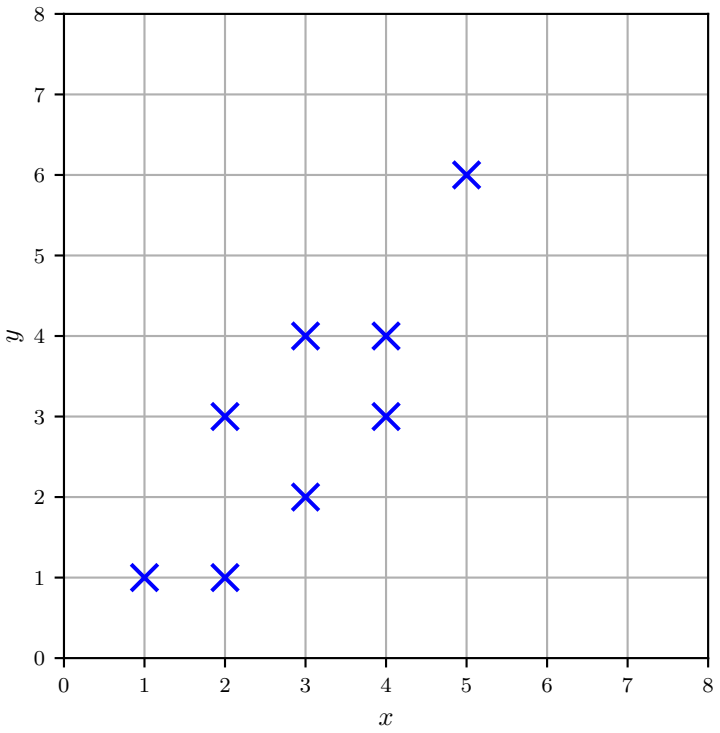
$$\hat{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i \quad \text{for} \quad i = 1, \ldots, N \tag{2}$$

where the directions that maximize the variance are given by sorting the eigenvectors with respect to the size of their associated eigenvalues $\lambda_{(1)} > \cdots > \lambda_{(N)}$. The first principal component is then $\mathbf{v}_{(1)}$, the second $\mathbf{v}_{(2)}$, and so on. They describe the variance in the data in descending order.

In the following, you are given a collection of $N = 8$ data points in a two-dimensional space:

$$X = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 2 & 3 \\ 3 & 2 \\ 3 & 4 \\ 4 & 3 \\ 4 & 4 \\ 5 & 6 \end{pmatrix}, \tag{3}$$

with $X \in \mathbb{R}^{N \times D}$, plotted below:



Tasks:

a) Center the data and compute its covariance matrix $\hat{C}$.

b) You are given the eigenvalues of the covariance matrix as $\lambda_1 = 4.23$ and $\lambda_2 = 0.34$. Compute the principal eigenvector $\mathbf{v}_1$ of the data covariance matrix and sketch it in the figure.

Hint: Use the fact that for an eigenvector $\mathbf{v}_i$, $\hat{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ and $\|\mathbf{v}_i\|_2 = 1$, and draw the eigenvector starting from the mean. Do not perform eigendecomposition.

c) If we project the data to a reduced order space using the principal component computed by the first eigenvector of the covariance matrix, how much of the variance of the data (in $\%$) do we expect to retain in the reduced order space? Do not project the data.

d) Use the principal component $\mathbf{v}_1$ computed in the previous questions to project the **centered** data in a reduced order space. Compute the variance of the data in the original 2D space, and the variance of the projected data. How much variance (in $\%$) is retained in the reduced order space? Does this number agree with the value computed in the previous question where we used the eigenvalues?

e) The computational cost of many data-driven algorithms (e.g. nearest neighbor search or classification) scales quadratically, i.e. $\mathcal{O}(N^2)$, or even cubically $\mathcal{O}(N^3)$ with the data dimension N. In the era of Big Data, many datasets have a large dimensionality, rendering the direct application of such algorithms computationally unfeasible. How can PCA be used to alleviate the issue? Keep your answer to five points at most.