Machine Learning for Autonomous Robots

Tutorial Feature Selection / PCA Example

Hans Hohenfeld

1 Introduction

In the tutorial on feature selection and dimensionality reduction we discussed a simple yet insightful example for the Principle Component Analysis (PCA) algorithm. This document provides all discussed details including some additional information especially on the mathematical background.

2 Principal Component Analysis

Main purpose of the PCA is to compress given data with regards to its dimensionality and decorrelate its dimensions while preserving as much inner structure of the data set as possible. Given a dataset $D := \{x_1, x_2, \ldots, x_n\}$, the PCA works by applying the following steps (for more details refer to the lecture material):

1. Organize D in a matrix $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix}$.
2. Center the data by subtracting the empirical mean $X \leftarrow X - \bar{X}$. Where $\bar{X} = \begin{pmatrix} \bar{x} \\ \bar{x} \\ \vdots \end{pmatrix}$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
3. Compute the empirical covariance matrix $\Sigma = \frac{1}{n-1} X^\intercal X$.
4. Compute the m eigenvalues λ_i and eigenvectors e_i of Σ by solving $\Sigma e_i = \lambda_i e_i$.
5. Organize the first $k \leq m$ eigenvectors in a projection matrix $W := (e_1 \cdots e_k)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ (The eigenvectors are sorted by their corresponding eigenvalues in descending order).
6. Project X by $Y \leftarrow XW$.

3 A simple PCA example

Given the following two-dimensional dataset with $n = 10$ data points we would like to perform a PCA:

	data									
x	2.5	0.5	2.2	1.9	3.1	2.3	2	1	1.5	1.1
y	2.4	0.7	2.9	2.2	3.0	2.7	1.6	1.1	1.6	0.9

Some first insight into the data can be gained by plotting it and visually inspecting its structure. Figure 1 shows the unmodified data and its already visible that on an estimated diagonal line from the bottom left to the top right the data has the greates variance.

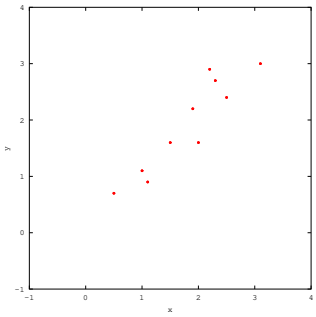

Figure 1: The original data

The empirical mean of the data set is $\bar{x} = (1.811.91)$ if substracted from the original data set we get:

$$X = \begin{pmatrix} 0.69 & 0.49 \\ -1.31 & -1.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ 0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -1.01 \end{pmatrix}$$

Plotting X shows, that the data has been centered arround the origin while its inner structure was kept intact.

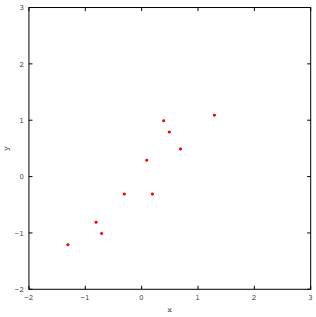

Figure 2: Centered data

We can get the data set's empirical covariance matrix Σ by computing $\Sigma = \frac{1}{n-1} X^\intercal X$ resulting in

$$\Sigma = \begin{pmatrix} 0.61656 & 0.61544 \\ 0.61544 & 0.71656 \end{pmatrix}.$$

From the covariance matrix we can read much of the information we already found through visual inspection of our data set. The diagonal elements show very similar variance in x and y direction with a

slightly bigger value for y. The positive off diagonal values (0.61544) indicate, that greater y values tend to occur for greater x values and vice versa.

We can now perform a **eigendecomposition** of the covariance matrix by solving

$$\Sigma e_i = \lambda_i e_i$$

Where e_i are the eigenvectors of Σ and λ_i the corresponding eigenvalues. First we need to transform our problem to a homogenous system of linear equations:

$$\begin{aligned} \Sigma e_i &= \lambda_i e_i \\ \Sigma e_i &= \lambda_i I e_i \\ \Sigma e_i - \lambda_i I e_i &= 0 \\ (\Sigma - \lambda_i I) e_i &= 0 \end{aligned}$$

where I is the identity matrix. This system has a nontrivial solution ($e_i \neq 0$) iff the coefficent matrix' dertminant is 0. Finding the eigenvalues λ_i therefore means calculating the nulls of the characteristic polynome of Σ:

$$\begin{aligned} \det(\Sigma - \lambda I) &= 0 \\ \det\left(\begin{pmatrix} 0.61656 & 0.61544 \\ 0.61544 & 0.71656 \end{pmatrix} - \lambda_i \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) &= 0 \\ \det\begin{pmatrix} 0.61656 - \lambda_i & 0.61544 \\ 0.61544 & 0.71656 - \lambda_i \end{pmatrix} &= 0 \\ \lambda_i^2 - 1.33312\lambda_i + 0.063036 &= 0 \end{aligned}$$

The above eqaution has two real valued solutions, $\lambda_1 = 1.284028$ and $\lambda_2 = 0.049083$ which are the eigenvalues of Σ. To get the eigenvectors e_i we now just need to enter each λ_i in the system of linear equations above and solve the resulting system (e. g., with Gaussian elimination). While this is possible for our two dimensional example and other small examples too, the problem quickly becomes intractable in practice. You would therefore want to use some implemented algorithm to find eigenvalues and eigenvectors of a matrix.

For our problem, we get the following two eigenvectors:

$$\begin{aligned} e_1 &= \begin{pmatrix} 0.67787 \\ 0.73518 \end{pmatrix} \\ e_2 &= \begin{pmatrix} -0.73518 \\ 0.67787 \end{pmatrix} \end{aligned}$$

Note: Usually there will indefinite many solutions, where the vectors for each eigenvalue have the same direction (or opposite) directions but different lengths. Please refer to the documentation for the implemented algorithm you use to learn about its return values.

Figure 3 shows the centered data with both eigenvectors. The blue line (e_1) confirms, what we assumed based on our visual inspection of the data. On this diagonal, the most information regarding the data is encoded (with regards to its variance). The second eigenvector e_2 is depicted in green. It has the second most value. Note that both vectors are orthogonal to each other and will server as new coordinate systems for our data, after we performed the projection.

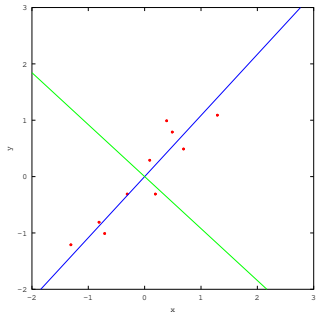

Figure 3: Centered data with the two eigenvectors e_1 (blue) and e_2 (green).

All that is left to do now, is to create a projection matrix $W = (e_1 e_2)$ and project our data onto our new coordinate system. Note that at this step we have to decide if we want to reduce our data's dimensionality. We can do so, by only selecting the first k eigenvectors (orderd by their corresponding eigenvalues in descending order) for W. As we only have two dimensions, we will not compress the data. We therefore get:

$$Y = XW = \begin{pmatrix} 0.827970 & -0.175115 \\ -1.777580 & 0.142857 \\ 0.992197 & 0.384375 \\ 0.274210 & 0.130417 \\ 1.675801 & -0.209498 \\ 0.912949 & 0.175282 \\ -0.099109 & -0.349825 \\ -1.144572 & 0.046417 \\ -0.438046 & 0.017765 \\ -1.223821 & -0.162675 \end{pmatrix}$$

And if we plot the projected data

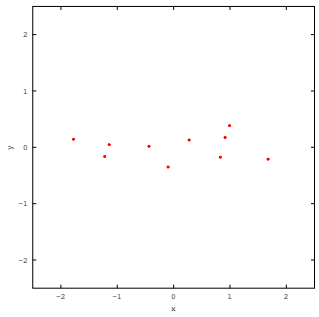

Figure 4: The projected data

we see, that it is centered arround the origin (its expected value is 0), has its greatest variance in x direction and its second greatest in y direction. We can confirm the decorrelation of the x and y dimension by checking Y's covariance matrix

$$\Sigma_Y = \begin{pmatrix} 1.284 & 0 \\ 0 & 0.049083 \end{pmatrix}$$

which supports our findings from visual inspection.

4 References

The example was taken from "A tutorial on Principal Component Analysis" by Lindsay I. Smith (2002), http://www.iro.umontreal.ca/~pift6080/H09/documents/papers/pca_tutorial.pdf. You will find more details on the mathematical background there. Otherwise please refer to the lecture material provided to you.