Ukrainian Catholic University

Applied Sciences Faculty

Data Science Master Programme

Kernel Principal Component Analysis and its Applications

Project report

Authors:

Irynei Baran

Hanna Pylieva

February 17, 2018

Abstract

Principal component analysis (PCA) is a popular tool for linear dimensionality reduction and feature extraction. Kernel PCA is the nonlinear form of PCA, which better exploits the complicated spatial structure of high-dimensional features. In this project, we first review the basic ideas of PCA and kernel PCA. Then we show some experimental results to compare the performance of kernel PCA and standard PCA for classification problems. We also provide an overview of PCA and kPCA applications.

revise Abstract after finishing

1 Introduction

In this section, we briefly review the principal component analysis method, it's applications and limitations.

rewrite the introduction

2 Principal Component Analysis

Principal component analysis, or PCA, is a mathematical procedure which is widely used for dimensionality reduction and feature selection. Those applications are achieved by projecting the data orthogonally onto a linear space with lower dimension, known as the principal subspace or feature space, such that the variance of projected data is maximal (Bishop , 2009).

Consider a data set X containing N observations of D features ($D < N$). In order to visualize data or diminish number of features for modeling we want to reduce dimensionality of feature space to $M < D$. Whereas we are interested in most influential features to minimize data losses, that is why our goal is to maximize variance of projected data. The directions on which the data is projected called principal components. They are orthogonal and form coordinate system of subspace M (see on Figure 1).
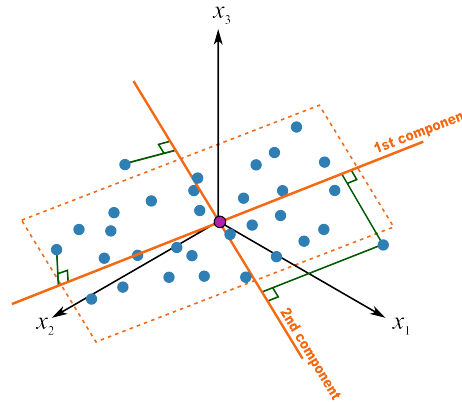

Figure 1: Principal components

2.1 Finding principal components

To begin with, consider the projection on M when $dim(M) = 1$. Let a unit vector $u_1 \in D$ be the direction of M. Then projection of an observation $x_n \in X$ onto M is $u_1^T x_n$ and the

1

variance of projected data is

$$\frac{1}{N} \sum_{n=1}^{N} \{u_1^T x_n - u_1^T \bar{x}\} = u_1^T S u_1 \tag{1}$$

where $\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_i$ is the mean of sample set and S is the covariance matrix of data set X.

Maximization of (1) is kept in a unit circle as we chose u_1 s.t. $||u_1|| = u_1^T u_1 = 1$. So we need to find maximum of the next Lagrange function:

$$L(X, \lambda_1) = u_1^T S u_1 + \lambda_1(1 - u_1^T u_1) \tag{2}$$

By setting the derivative with respect to u_1 equal to zero we find that in stationary point u_1 needs to be an eigenvector of S:

$$S u_1 = \lambda_1 u_1 \tag{3}$$

Now when we left-multiply by u_1^T and make use of $u_1^T u_1 = 1$ we find out that the variance is given by

$$u_1^T S u_1 = \lambda_1 \tag{4}$$

and so the variance will be a maximum when we set u_1 equal to the eigenvector having the largest eigenvalue λ_1. This eigenvector id called the first principal component (Bishop , 2009).

Next principal components can be found following the same procedure and choosing each new direction such that it maximizes the projected variance amongst all possible directions orthogonal to those already considered.

2.2   Limitations of standard PCA

Although PCA is very useful from practical prospective it has some limitations.

1. Assumptions of linear dependency. PCA projects data orthogonally to reduce dimensionality. This works only if data has linear dependent variables: $y = kx + \epsilon$. Otherwise the method doesn't identify the direction where projected data has highest possible variance which can be seen on Figure 2 .
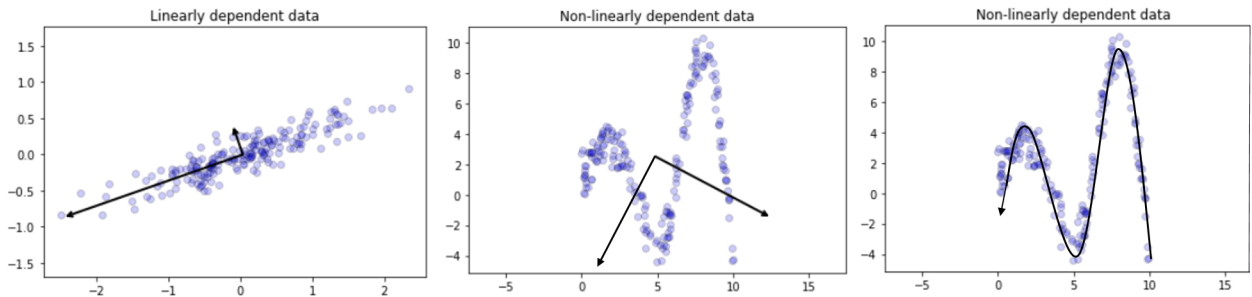

Figure 2: Demonstration of PCA's low performance for non-linearly dependent data

do we need to provide formula of covariance matrix explicitly? I don't feel so..
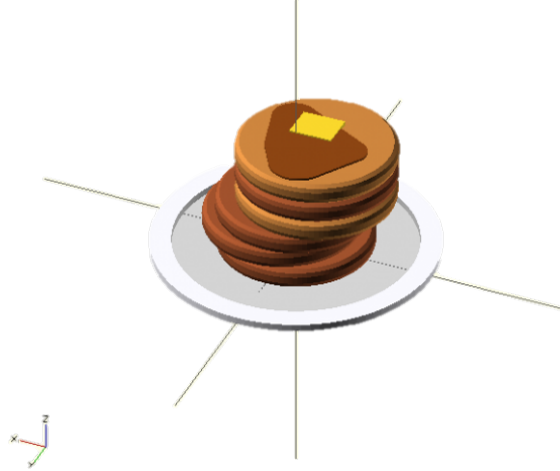
Figure 3: Image from 3D Parametric Pancakes

2. Spread maximization. PCA searches for a subspace where projected data has the maximal spread. However, this is not always the best way to represent data in lower-dimension subspace. A common example is the task of separating and counting pancakes from an image (see Figure 3). Important information for the task (number of pancakes) is located along Z axis which has the lowest variance. So Z axis will be the last principal component identified by PCA which is not what required.

3. Principal components are hard to interpret. PCA reveals implicit dependencies in data which is non-trivial to interpret in case of high-dimensional space reduction. The leverage for this issue is deep understanding of domain.

4. Orthogonality assumptions. PCA comes up with orthogonal principal components, which do not overlap in the space. For some tasks such functionality will produce wrong results (see Figure 4).
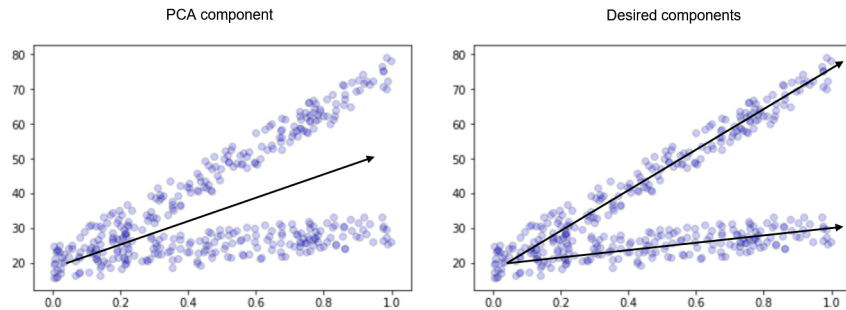

Figure 4: Demonstration of PCA's low performance for non-linearly dependent data

5. Outliers and missing data sensitivity. We introduced PCA here as eigenvalue decomposition of a data covariance matrix. The latter is highly sensitive to sample outliers and corrupted data (also referred as intra-sample outliers) (Shlens , 2014). This raises two issues:

3

(a) PCA is a linear combinations of all input variables whereas in case of corrupted or missing entries presence we need to exclude some incomplete variables.

(b) in some contexts, outliers can be difficult to identify. In this case we are unable to remove them in order to get better performance from PCA. Whereas there is a way to overcome the issue by introducing robustness to the algorithm. This will be considered beneath.

3 Modifications of PCA

Despite the limitations mentioned, PCA is still a powerful method for dimensionality reduction. That is why numerous modifications were developed to overcome its drawbacks and broaden scope of applications. We will provide succinct description of the most common extensions of PCA.

3.1 Robust PCA

Gross errors in observations are now ubiquitous in modern applications such as image processing, web data analysis, and bioinformatics, where some measurements may be arbitrarily corrupted (due to occlusions, malicious tampering, or sensor failures) or simply irrelevant to the low-dimensional structure we seek to identify with PCA.The best algorithm which deals with data corruption is Robust PCA (RPCA).

Suppose the data under study can naturally be decomposed into low-rank (L_0) and sparse (S_0) components[1]:

$$M = L_0 + S_0 \tag{5}$$

RPCA allows to recover a low-rank matrix L_0 from highly corrupted measurements in M. Classical PCA works good when the noise term N_0 in M is small. In contrast to this S_0 can have arbitrarily large magnitude, and their support is assumed to be sparse but unknown (Cardes , 2009).

3.2 Multilinear PCA

Suppose we need to reduce dimensionality not of a simple matrix $M \in R^2$ but of a n-way array, i.e. a cube or hyper-cube of numbers, also informally referred to as a "data tensor". Common examples of tensor are 2-D/3-D images and video sequences. To deal with tensors PCA is generalized to multilinear PCA (MPCA). MPCA performs feature extraction by determining a multilinear projection that captures most of the original tensorial input variation. The solution is iterative in nature and it proceeds by decomposing the original problem to a series of multiple projection subproblems (Lu , 2008).

MPCA is applied to 3-D object recognition tasks (Sahambi , 2003) in machine vision, medical image analysis, space-time analysis of video sequences for activity recognition (Green , 2004) in human-computer interaction, etc. MPCA is further extended to uncorrelated MPCA, non-negative MPCA and robust MPCA.

---

[1]This is not a synthetic requirement and can be done in a number of applications, refer to (Cardes , 2009) for examples.

References

Bishop, Christopher M. *Pattern Recognition and Machine Learning.* Springer, Cambridge, U.K., 2006.

Shlens, Jonathon. *A Tutorial on Principal Component Analysis.*Google Research, Mountain View, USA, 2014.

Candes, Emmanuel J.; Li, Xiaodong ; Ma, Yi; Wright,John. *Robust Principal Component Analysis?* J.ACM, 58(11):1-37, 2009

Lu, Haiping; Plataniotis, Konstantinos N.; Venetsanopoulos, Anastasios N. *MPCA: Multilinear Principal Component Analysis of Tensor Objects.* IEEE Transactions on Neural Networks, 19(1):18-39, 2008

Sahambi, Harkirat S.; Khorasani, Khakhayar. *A neural-network appearance-based 3-D object recognition using independent component analysis* IEEE Transactions on Neural Networks, 14(1): 138-149, 2003.

Green, Richard D.; Guan, Ling. *Quantifying and recognizing human movement patterns from monocular video images - Part II: Applications to biometrics.* IEEE Transactions on Circuits and Systems for Video Technology, 14(2): 191-198, 2004.