

KBLMM - Homework 4

Kernel Principal Component Analysis

Mikołaj Małkiński

28 October, 2019

Contents

1	Introduction	1
2	Principal Component Analysis	2
3	Kernel Principal Component Analysis	2

1 Introduction

Principal Component Analysis is a dimensionality reduction technique. Its main goal is to reduce the number of features of a given dataset in such a way that most of the original variance of the dataset is explained in the lower dimensional space. One of its use cases is to project the data onto a 2D or 3D plane where it can be easily visualised. Also, PCA can be useful as a preprocessing technique before passing data to a machine learning model, because usually they generalise better when dealing with fewer number of features. However, this will work only if the chosen principal components are able to explain most of the variance in a dataset.

PCA is a linear method, therefore oftentimes it may be not enough. That is why a Kernel Principal Component Analysis was introduced, which is able to deal with nonlinearity by using different kernel functions. The idea of kPCA is to firstly map the data into a higher-dimensional Hilbert space, in which the dataset might become linearly separable.

The goal of this work is to showcase the difference between PCA and kPCA when dealing with nonlinear data. The datasets were generated with the help of *clusterSim* package and are shown in the Figure 1.

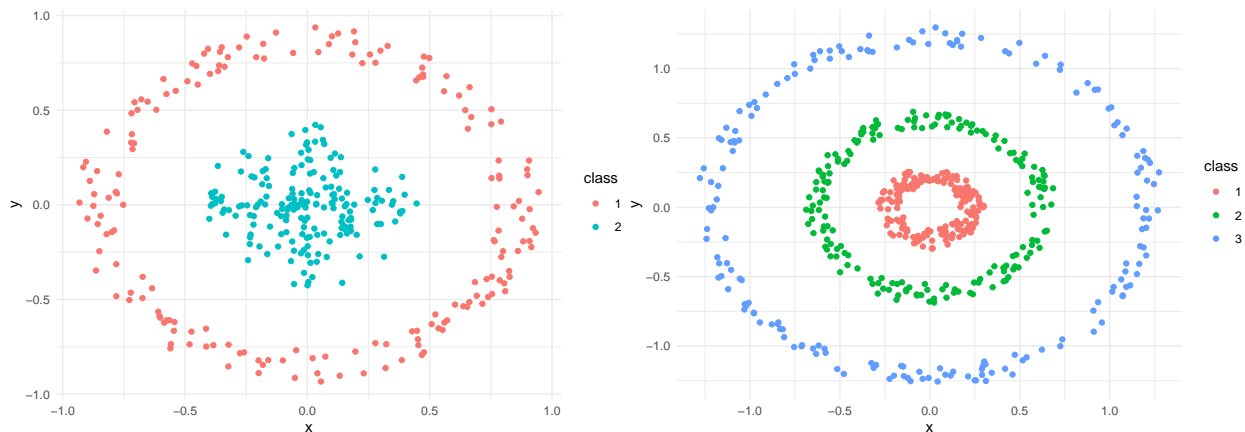


Figure 1: Bulls eye (left) and 3 circles (right) datasets

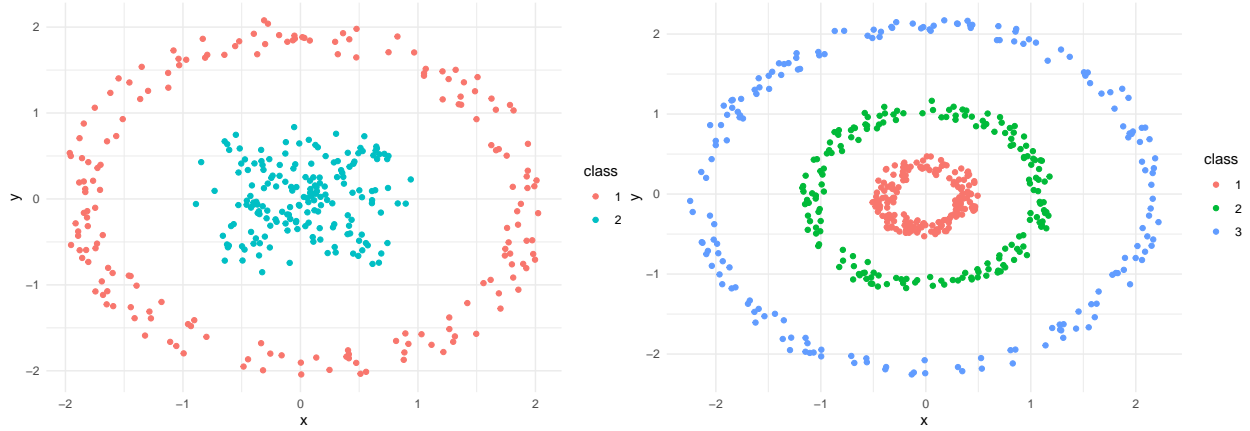


Figure 2: Projection of datasets using 2 first principal components from PCA

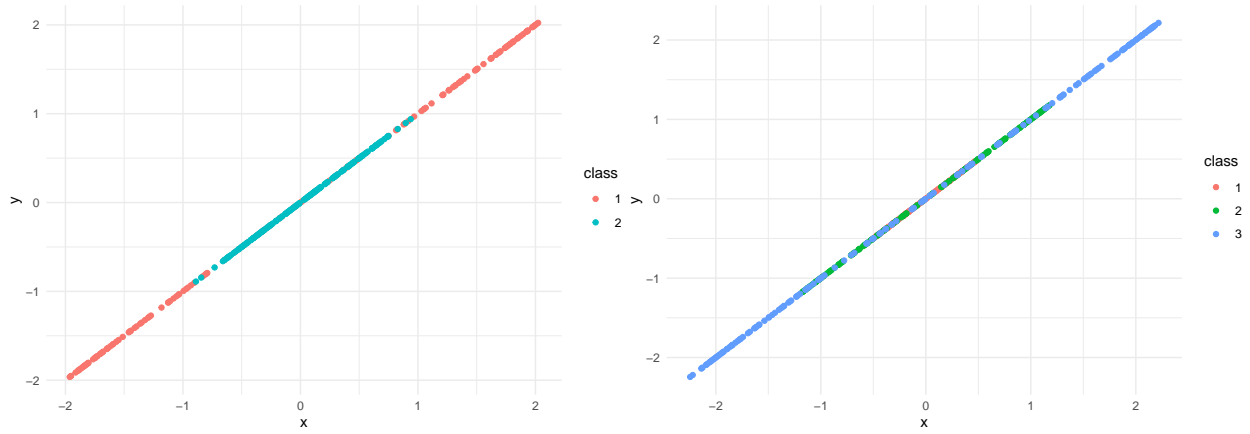


Figure 3: Projection of datasets using first principal component from PCA

2 Principal Component Analysis

Firstly, we perform PCA for both datasets. Then we project it onto a new space using selected number of principal components. Figure 2 shows the projection for both datasets using 2 first principal components. It is visible that the projection didn't change the data too much, it only rotated it. Furthermore, Figure 3 shows the projection of this data using only first principal component. Clearly, the data isn't linearly separable and PCA didn't give useful results when working with a nonlinear dataset.

3 Kernel Principal Component Analysis

Now the same experiments are performed, but this time kernel PCA is used with the RBF kernel (with $\sigma = 3$). The kernel together with its hyperparameter was chosen by manual experimentation - several kernels and their hyperparameter configurations were tried and the best performing one was selected. Figure 4 shows the projection for both datasets using 2 first principal components. This time, the projected data can be clearly separated with a line. Moreover, it is enough to project the data using only the first principal component as shown in the Figure 5. This experiments show that kernel PCA is able to deal with nonlinearities in datasets by projecting data into different space, where linear separation is possible.

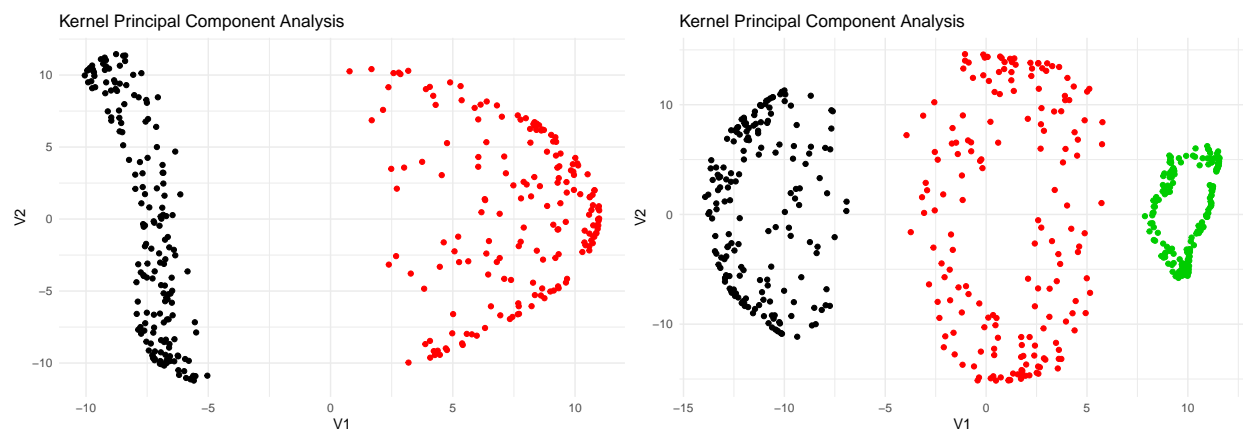


Figure 4: Projection of datasets using 2 first principal components from kernel PCA, using RBF kernel

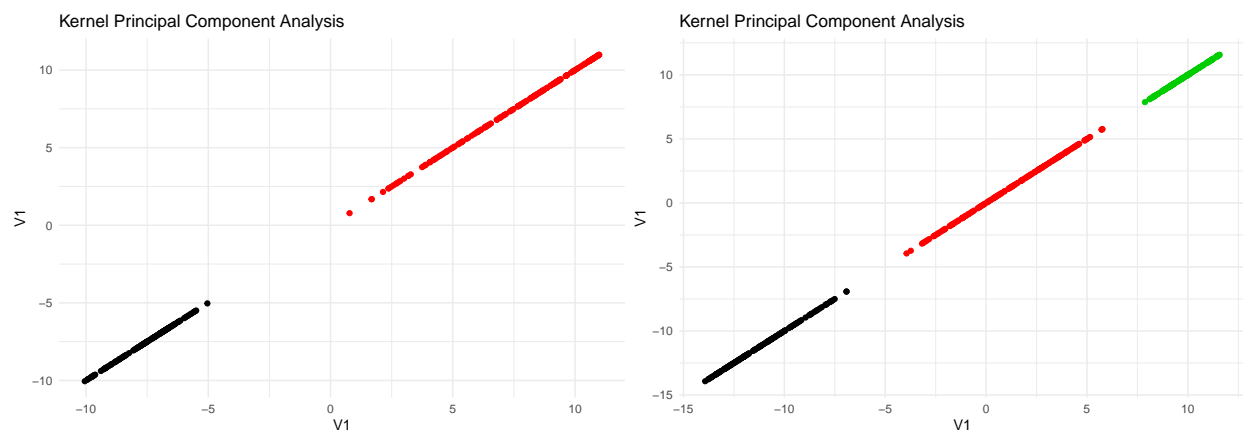


Figure 5: Projection of datasets using first principal component from kernel PCA, using RBF kernel