

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 13, 2018



APPLIED
SCIENCES
FACULTY ●

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.

1 Introduction

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

1.1 Principal Component Analysis

Principal component analysis, or PCA, is a technique that is widely used for applications such as dimensionality reduction, lossy data compression, feature extraction, and data visualization. PCA can be defined as the orthogonal projection of the data onto a lower dimensional linear space, known as the principal subspace, such that the variance of the projected data is maximized (Bishop, 2006).

2 Examples

2.1 Comments

Comments can be added to the margins of the document using the `todo` command, as shown in the example on the right. You can also add inline comments too:

This is an inline comment.

Here's
a com-
ment
in the
mar-
gin!

2.2 Tables and Figures

Use the `table` and `tabular` commands for basic tables — see Table 1, for example. You can upload a figure (JPEG, PNG or PDF) using the `files`



Figure 1: This is a figure caption.

Item	Quantity
Widgets	42
Gadgets	13

Table 1: An example table.

menu. To include it in your document, use the `includegraphics` command as in the code for Figure 1 below.

2.3 Mathematics

L^AT_EX is great at typesetting mathematics. Let X_1, X_2, \dots, X_n be a sequence of independent and identically distributed random variables with $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2 < \infty$, and let

$$S_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_i^n X_i$$

denote their mean. Then as n approaches infinity, the random variables $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$.

2.4 Lists

You can make lists with automatic numbering ...

1. Like this,
2. and like this.

...or bullet points ...

- Like this,
- and like this.

We hope you find write^LA_TE_X useful, and please let us know if you have any feedback using the help menu above.

References

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