PCA

Principal Component Analysis

Wednesday 16th November, 2016

Parabhat Soni

MTech CSE

OVERVIEW

1. Introduction
2. Pre-requisites
3. Derivation
4. PCA - Steps to be followed
5. Useful properties of PCA
6. gPCA - Generalized Principal Component Analysis

INTRODUCTION

WHAT IS PCA?

Central idea of principal component analysis:

- reduce of dimensionality of a data set consisting of large number of interrelated variables
- retaining as much as possible of the variation present in the data set.
- This is achieved by transforming to a new set of variables, principal components(PCs)
- Pcs are uncorrelated and ordered so that the first few retains the most of the variation present.

WHAT IS PCA?

PCA finds a linear projection of high dimensional data into a lower dimensional subspace such as:

- The variance retained is maximized.(??)

PRE-REQUISITES

PRE REQUISITES

Some of the pre requisites:

1. Eigen vectors
   1.1 If A is a square matrix then a non zero vector v is an eigenvector of A if there is a scalar quantity such that

$$Av = \lambda v$$

DERIVATION

PCA - DERIVATION

Given a zero-mean random variable

$$X \epsilon R^D,$$

the first principal component of X is defined as

$$v_1 = argmax_{v1} var(Xv_1)$$

$$Subject\ to : v_1^T v_1 = 1$$

Rewrite in terms of the covariance matrix:

$$var(Xv_1) = \frac{1}{N-1}(Xv_1)^T(Xv_1) = v_1^T C v_1$$

Solved via constrained optimisation using Lagranges method:

$$L(v_1, \lambda_1) = v_1^T C v_1 + \lambda_1(1 - v_1^T v_1)$$

PCA - DERIVATION CONTD...

Constrained Optimisation:

$$L(v_1, \lambda_1) = v_1^T C v_1 + \lambda_1(1 - v_1^T v_1)$$

Gradient with respect to v:

$$\frac{dL(v_1, \lambda_1)}{dv_1} = 2Cv_1 - 2\lambda_1 v1 \Rightarrow Cv_1 = \lambda_! v_1$$

This is the eigenvector problem!

Multiply by v1T:

$$\lambda_1 = v_1^T C v_1$$

The projection variance is the eigenvalue

And further we can find more PCs.

PCA - STEPS TO BE FOLLOWED

PCA - STEPS TO BE FOLLOWED

PCA Steps: Transform an N x d matrix into an N × m matrix :

- Centralized the data (subtract the mean).
- Calculate the × covariance matrix:

$$C = \frac{1}{N-1} X^T X$$

- Calculate the eigenvectors of the covariance matrix.
- Select m eigenvectors that correspond to the largest m eigenvalues to be the new basis.

USEFUL PROPERTIES OF PCA

USEFUL PROPERTIES OF PCA

PCA Steps: Transform an N x d matrix into an N × m matrix :

- Centralized the data (subtract the mean).
- Calculate the × covariance matrix:

$$C = \frac{1}{N-1} X^T X$$

- Calculate the eigenvectors of the covariance matrix.
- Select m eigenvectors that correspond to the largest m eigenvalues to be the new basis.

GPCA - GENERALIZED PRINCIPAL COMPONENT ANALYSIS

THEME OPTIONS

This theme comes with some options to change it's appearance.

Option	Description
`nosectionpages`	Section pages will be supressed.

PRESENTATION STRUCTURE

A section page will be generated and the section name included in the presentation header for each section of the presentation with the current section being emphasized. If you include subsections in your presentation, then a small block will appear under the section name in the header for each frame. Once a frame has been viewed it will turn green.

It's worth noting that a frame can make up multiple slides.

```
\section{Main Section}
\subsection{Main Subsection}
\begin{frame}
\frametitle{Presentation Stucture}
% Frame Contents Here
\end{frame}
```

TABLE OF CONTENTS

Include a listing of the presentation's sections

```
\maketitle
```

For those longer presentations - keep the table of contents compact.

```
\begin{frame}{Overview}
    \tableofcontents[hideallsubsections]
\end{frame}
```

QUOTATIONS

At any time you can highlight text by using the `\alert` definition:

- **This is super important!**

The sthlm presentation has both `\lstinline` and `\lstinline!` definitions for quoting text.

>This text has been quoted<

»This text has been double quoted«

NOTES

There are a couple different ways in which you can present your presentation.

- Splitshow (Mac OS X)
  `https://code.google.com/p/splitshow/`
- pdf-presenter (Windows)
  `https://code.google.com/p/pdf-presenter/`

NOTES

There are a couple different ways in which you can present your presentation.

- Splitshow (Mac OS X)
  `https://code.google.com/p/splitshow/`
- pdf-presenter (Windows)
  `https://code.google.com/p/pdf-presenter/`

Let me include a note for this particular slide.

- Splitshow (Mac OS X)
  `https://code.google.com/p/splitshow/`
- pdf-presenter (Windows)
  `https://code.google.com/p/pdf-presenter/`

MULTIPLE COLUMNS

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

- Point 1
- Point 2

REFERENCES

Alan V. Oppenheim
Discrete-Time Signal Processing
Prentice Hall Press, 2009

European Broadcasting Union
Specification of the Broadcast Wave Format (BWF)
2011

ABOUT

This sthlm beamer theme is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.
If you have any questions or comments

- sayhi@hendryolson.com