PCA Revealed

Part 6: Minimization Approach

Gaston **S**anchez

August 2014

Readme

License:

You are free to:

Under the following conditions:

Presentation

About

In these slides we'll cover PCA from a minimization problem perspective

Working Principle

The underlying mechanism behind this approach is the **Least Squares** (LS) principle

Reminder

PCA

Principal Components Analysis (PCA) allows us to study and explore a set of quantitative variables measured on a set of objects.

Core Idea

With PCA we seek to reduce the dimensionality (reduce the number of variables) of a data set while retaining as much as possible of the variation present in the data.

Data Modeling Perspective

Data Structure

Data

The data structure for PCA is in tabular format, which can be mathematically handled as a **matrix** $\mathbf{X}$:

$$\mathbf{X}_{n,p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

- n objects in the rows
- p quantitative variables in the columns

Data Considerations

Variables

We will denote the p variables in $\mathbf{X}$ by $X_1, X_2, \dots, X_p$

Mean centered

For convenience (to make computations easier and notation simpler) we will assume that the data is centered

$$\bar{X}_j = \sum_{i=1}^{n} x_{ij} = 0$$

(i.e. *centered*: variables with mean $= 0$)

Overall Idea

The main idea is to look at the data from a modeling perspective:

$$\text{Data} = \text{Fit} + \text{Error}$$

$$\text{Data} = \text{Structure} + \text{Residual}$$

Data Modeling (con't)

Data = Structure + Residual

Model Fit

Data Fit

Given a dataset $\mathbf{X}$, we want to find principal components $\mathbf{Z}$ and loadings $\mathbf{P}$ such that $\mathbf{ZP}'$ "optimally" fits $\mathbf{X}$

In other words

We want $\mathbf{Z}$ and $\mathbf{P}$ such that the model

$$\mathbf{X} = \mathbf{ZP}' + \mathbf{Error}$$

has the smallest amount of $\mathbf{Error}$ possible.

Some Comments

Model

Let's take a closer look at the model expression

$$\mathbf{X} = \mathbf{Z}\mathbf{P}' + \mathbf{Error}$$

Remarks

- Initially, both $\mathbf{Z}$ and $\mathbf{P}$ are unknown
- If we knew $\mathbf{Z}$ then we would need to find $\mathbf{P}$
- If we knew $\mathbf{P}$ then we would need to find $\mathbf{Z}$
- This is a **bilinear model**, because of the product term $\mathbf{Z}\mathbf{P}'$

Some Comments

Model

Let's keep looking at the model expression

$$\mathbf{X} = \mathbf{Z}\mathbf{P}' + \mathbf{Error}$$

Remarks

With no extra requirements or additional conditions, this model is very open (and not very useful). To enclose it further we need to add a very important specification:

$$\mathbf{Z} = \mathbf{X}\mathbf{W}$$

that is, we want to express the principal components in $\mathbf{Z}$ as linear combinations of the variables in $\mathbf{X}$

Model Fit (con't)

Optimal Fit?

In order to propose a solution we need to be more specific about the meaning of the "optimal fit" notion.

Least Squares

Among the several ways in which we could look for minimal $Error$, the chosen one in PCA (as in many other multivariate methods) is that of: minimal sum of squares error

In other words

We want $\mathbf{Z}$ and $\mathbf{P}$ such that the model

$$\underset{Z,P}{minimize} \; \|\mathbf{X} - \mathbf{ZP'}\|^2$$

the square norm of the difference between $\mathbf{X}$ and $\mathbf{ZP'}$ is as small as possible

Optimization Criterion

Loss function

Overall Idea

The loss function to be minimized is the loss information due to representing the variables by a small number of components

$$\underset{Z,P}{minimize} \; \|\mathbf{X} - \mathbf{ZP'}\|^2$$

(i.e. the difference between the variables and the component scores weighted by the component loadings)

Equivalent Expressions

Model

$$\mathbf{X} = \mathbf{Z}\mathbf{P}' + \mathbf{Error}$$

Loss Function forms

The proposed solution consists of minimizing a least-squares loss function

$$\begin{aligned} L(X, Z, P) &= \|\mathbf{X} - \mathbf{Z}\mathbf{P}'\|^2 && (1) \\ &= tr(\mathbf{X} - \mathbf{Z}\mathbf{P}')'(\mathbf{X} - \mathbf{Z}\mathbf{P}') && (2) \\ &= \frac{1}{n} \sum_{j=1}^{k} tr(X_j - \mathbf{Z}p_j)'(X_j - \mathbf{Z}p_j) && (3) \end{aligned}$$

Restrictions

Minimization Criterion

$$\underset{Z,P}{minimize} \; \|\mathbf{X} - \mathbf{ZP'}\|^2$$

where $\mathbf{Z} = \mathbf{XW}$

Subject to

$$\frac{1}{n}\mathbf{Z'Z} = \frac{1}{n}\mathbf{W'X'XW} = \mathbf{I}$$

Minimization

Minimization Criterion

To minimize the loss function, we start by eliminating $\mathbf{P}$. It can be shown that $\mathbf{P}' = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$.

Using the previous expression for $\mathbf{P}$ leaves us with minimizing:

$$L(\mathbf{W}) = \|\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{W}'\mathbf{R}\|^2$$

where $\mathbf{R} = \frac{1}{n}\mathbf{X}'\mathbf{X}$

subject to

$$\frac{1}{n}\mathbf{W}'\mathbf{X}'\mathbf{X}\mathbf{W} = \mathbf{I}$$

Finding PCs

LS Solution

Minimization Criterion

$$L(\mathbf{W}) = \|\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{W}'\mathbf{R}\|^2$$

where $\mathbf{R} = \frac{1}{n}\mathbf{X}'\mathbf{X}$

subject to

$$\frac{1}{n}\mathbf{W}'\mathbf{X}'\mathbf{X}\mathbf{W} = \mathbf{I}$$

LS Solution

Minimization Criterion

The solution is obtained with the Eigenvalue Decomposition (EVD) of $\mathbf{R}$:

$$\mathbf{R} = \mathbf{U \Lambda U}'$$

Weight Coefficients

The matrix of weights $\mathbf{W}$ to get the PCs is given by:

$$\mathbf{W} = \mathbf{R}^{-1/2}\mathbf{U} = \mathbf{U \Lambda}^{-1/2}$$

So far we've seen that ...

Looking for PCs

Overall Idea

Given a set of p variables $X_1, X_2, \dots, X_p$, we want to obtain new k variables $Z_1, Z_2, \dots, Z_k$, called the **Principal Components** (PCs)

PCs in matrix format

$$\mathbf{Z}_{n,k} = \mathbf{X}_{n,p}\mathbf{W}_{p,k}$$

PCs are just linear combinations of $\mathbf{X}$

Looking for PCs

PC as linear combinations

$$
\begin{array}{rcl}
\mathsf{PC}_1 \longrightarrow & & Z_1 = w_{11}X_1 + w_{12}X_2 + \cdots + w_{1p}X_p \\
\mathsf{PC}_2 \longrightarrow & & Z_2 = w_{21}X_1 + w_{22}X_2 + \cdots + w_{2p}X_p \\
\vdots & & \vdots \\
\mathsf{PC}_k \longrightarrow & & Z_k = w_{k1}X_1 + w_{k2}X_2 + \cdots + w_{kp}X_p
\end{array}
$$

(i.e. linear combination = weighted sum)

Model fit illustration

Data = Structure + Residual

PCA Model

PCA model

Formally, PCA involves finding scores and loadings such that the data can be expressed as a product of two matrices:

$$\mathbf{X}_{n,p} = \mathbf{Z}_{n,k}\mathbf{P'}_{k,p}$$

where $\mathbf{Z}$ is the matrix of PCs or *scores*, and $\mathbf{P}$ is the matrix of *loadings*

Ideally

We expect k to be much more smaller than p so we get a data reduction without losing too much information.

PCA and Data Decomposition

Computation of all PCs

We can obtain as many PCs as the rank of $\mathbf{X}$ (i.e. $k = rank(\mathbf{X})$)

$$\mathbf{X}_{n,p} = \mathbf{Z}_{n,k}\mathbf{P'}_{k,p}$$

Keeping just a few PCs

But usually we will only retain just a few PCs (i.e. $k \ll p$)

$$\mathbf{X}_{n,p} \approx \mathbf{Z}_{n,k}\mathbf{P'}_{k,p} + Residual$$

(just a few PCs will *optimally* summarize the main structure of the data)