PCA Revealed

Part 9: Appendix

Gaston **S**anchez

August 2014

Readme

License:

This document is licensed under a

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

You are free to:

Share — copy and redistribute the material

Adapt — rebuild and transform the material

Under the following conditions:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

NonCommercial — You may not use this work for commercial purposes.

ShareAlike — If you remix, transform, or build upon this work, you must distribute your contributions under the same license to this one.

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

A bit of Algebra beihnd PCA

Capturing variation with PCs

Variation

Looking for PCs that *capture most of the variation in the data* implies —in statistical terms— that we want to obtain **PCs with maximum variance**

In other words

We look for vectors of weights $\mathbf{w_j} = \{w_{j1}, w_{j2}, \dots, w_{jp}\}$ such that each component $Z_j = \mathbf{X}\mathbf{w_j}$ has maximum variance (for $j = 1, \dots, k$)

Algebraic Formulation

More formally

We want to find a vector $\mathbf{w_j}$ such that

$$\max_{\mathbf{w_j}} \; var(Z_j = \mathbf{X}\mathbf{w_j})$$

that is

$$\max_{\mathbf{w_j}} \; \mathbf{w_j'}\mathbf{X}'\mathbf{X}\mathbf{w_j}$$

Note that

- $\mathbf{X}'\mathbf{X}$ is the variance-covariance matrix
- Without constraints, the previous expression is unbounded

Maximization Constraints

Usefull Restriction

To get a feasible solution we need to impose the restriction that $\mathbf{w_j}$ is of unit norm: $\|\mathbf{w_j}\| = 1 \;\Rightarrow\; \mathbf{w_j'w_j} = 1$

Criterion to be maximized

If we denote $\mathbf{S} = \mathbf{X'X}$ then the criterion to be maximized is:

$$\max_{\mathbf{w_j}} \; \mathbf{w_j'Sw_j}$$

subject to $\quad \mathbf{w_j'w_j} = 1 \quad$ and $\quad Z_j' Z_h = 0 \quad (j \neq h)$

Pay Attention!

$$\mathbf{w_j'Sw_j}$$

This expression is of extreme importance. Why?

- It is a quadratic form
- $\mathbf{S}$ is a semi-positive definite matrix
- $\mathbf{S}$ has non-negative real eigenvalues

Finding all PCs

All PCs can be found simultaneously by **diagonalizing** $\mathbf{S}$

Diagonalizing $\mathbf{S}$ involves expressing it as the product:

$$\mathbf{S} = \mathbf{W}\mathbf{D}\mathbf{W}'$$

- $\mathbf{D}$ is a diagonal matrix
- the elements in the diagonal of $\mathbf{D}$ are the eigenvalues of $\mathbf{S}$
- the columns of $\mathbf{W}$ are orthonormal: $\mathbf{W}'\mathbf{W} = \mathbf{I}$
- the columns of $\mathbf{W}$ are the eigenvectors of $\mathbf{S}$
- $\mathbf{W}' = \mathbf{W}^{-1}$

Diagonalization of $\mathbf{S}$

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1p} \\ w_{21} & w_{22} & \cdots & w_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ w_{p1} & w_{p2} & \cdots & w_{pp} \end{pmatrix}$$

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}$$

$$\mathbf{S} = \mathbf{W}\mathbf{D}\mathbf{W}'$$

Diagonalization of $\mathbf{S}$

$$\mathbf{S} = \mathbf{W}\mathbf{D}\mathbf{W}' =$$

$$\begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1p} \\ w_{21} & w_{22} & \cdots & w_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ w_{p1} & w_{p2} & \cdots & w_{pp} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} \begin{pmatrix} w_{11} & w_{21} & \cdots & w_{p1} \\ w_{12} & w_{22} & \cdots & w_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1p} & w_{2p} & \cdots & w_{pp} \end{pmatrix}$$

Diagonalization and Eigenvalue Decomposition

Diagonalization Reminder

Diagonalization?

A symmetric matrix $\mathbf{S}$ is **orthogonally diagonalizable** if and only if there is an orthogonal matrix $\mathbf{W}$ and a diagonal matrix $\mathbf{D}$ such that

$$\mathbf{S} = \mathbf{W}\mathbf{D}\mathbf{W}'$$

$$\mathbf{D} = \mathbf{W}'\mathbf{S}\mathbf{W}$$

Note that

- $\mathbf{W}' = \mathbf{W}^{-1}$
- the columns of $\mathbf{W}$ are orthonormal: $\mathbf{W}'\mathbf{W} = \mathbf{I}$
- the elements in the diagonal of $\mathbf{D}$ are the eigenvalues of $\mathbf{S}$
- the columns of $\mathbf{W}$ are the eigenvectors of $\mathbf{S}$

Spectral Theorem and Diagonalization

Eigenvalue Decomposition (EVD)

Diagonalizing a symmetric matrix is nothing more than obtaining its **eigenvalue decomposition** (a.k.a. spectral decomposition)

Spectral Decomposition of Symmetric Matrices

A $p \times p$ symmetric matrix $\mathbf{S}$ has the following properties:

- $\mathbf{S}$ has p real eigenvalues (counting multiplicites)
- the eigenvectors corresponding to different eigenvalues are orthogonal
- $\mathbf{S}$ is orthogonally diagonalizable ($\mathbf{S} = \mathbf{WDW}'$)
- the set of eigenvalues of $\mathbf{S}$ is called the **spectrum** of $\mathbf{S}$

Diagonalization and Change of Variable

Change of variable

The notion of *diagonalization* involves the idea of **change of variable**

From $\mathbf{v}$ to $\mathbf{u}$ and viceversa

If $\mathbf{v}$ is a vector (i.e. variable) in R^p, we can change $\mathbf{v}$ into $\mathbf{u}$ using the orthogonal matrix $\mathbf{W}$:

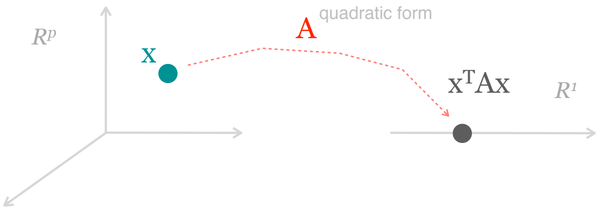
$$\mathbf{u} = \mathbf{W}\mathbf{v} \;\text{ or }\; \mathbf{v} = \mathbf{W}^{-1}\mathbf{u}$$

Quadratic Form

A very important concept to keep in mind is that of **quadratic form**. Basically, a quadratic form is a function f defined on R^p of the form:

$$f(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$$

where $\mathbf{A}$ is a $p \times p$ symmetric matrix



Role played by Quadratic Forms

Change of variable

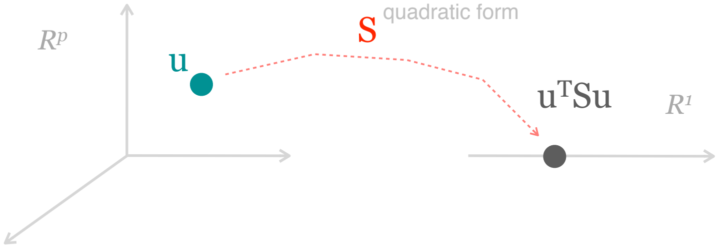
We can associate a change of variable $\mathbf{u} = \mathbf{W}\mathbf{v}$ to quadratic forms:

- $\mathbf{u}'\mathbf{S}\mathbf{u} \longrightarrow (\mathbf{W}\mathbf{v})'\mathbf{S}(\mathbf{W}\mathbf{v})$
- $\mathbf{v}'\mathbf{D}\mathbf{v} \longrightarrow (\mathbf{W}'\mathbf{u})'\mathbf{D}(\mathbf{W}'\mathbf{u})$

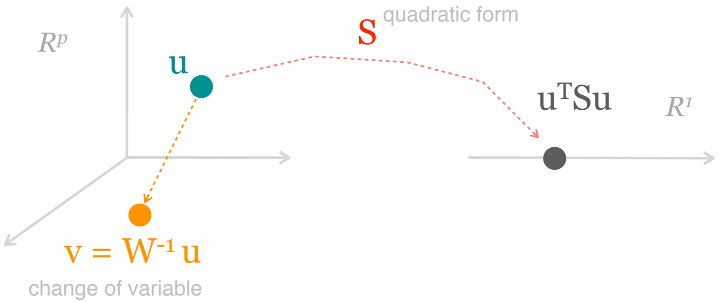
Relation between $\mathbf{u}'\mathbf{S}\mathbf{u}$ and $\mathbf{v}'\mathbf{D}\mathbf{v}$

$$\mathbf{u}'\mathbf{S}\mathbf{u} = (\mathbf{W}\mathbf{v})'\mathbf{S}\mathbf{W}\mathbf{v} = \mathbf{v}'\mathbf{W}'\mathbf{S}\mathbf{W}\mathbf{v} = \mathbf{v}'(\mathbf{W}'\mathbf{S}\mathbf{W})\mathbf{v} = \mathbf{v}'\mathbf{D}\mathbf{v}$$

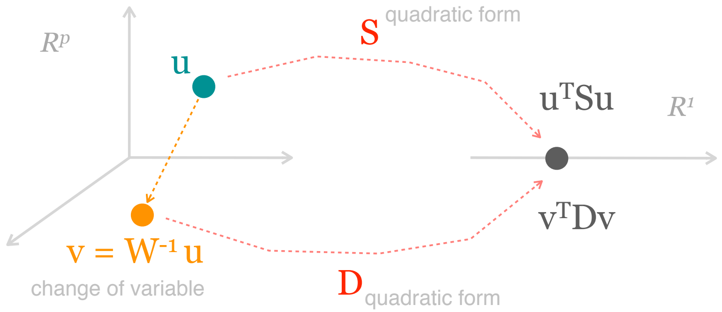
Quadratic Forms and Change of Variable



Quadratic Forms and Change of Variable



Quadratic Forms and Change of Variable



Principal Axes Theorem

Hyper-Ellipsoids

Intimately related with quadratic forms, we have expressions of the form

$$\mathbf{u}'\mathbf{S}\mathbf{u} = c$$

(with c constant) that define a very specific geometric object; **hyper-ellipsoids**

Principal Axes

The columns of $\mathbf{W}$ are called the **principal axes** of the quadratic form $\mathbf{u}'\mathbf{S}\mathbf{u}$

What does a change of variable look like?

PCA and EVD

The PCA solution can be obtained with an Eigenvalue Decomposition of the matrix $\mathbf{S} = \mathbf{X}'\mathbf{X}$

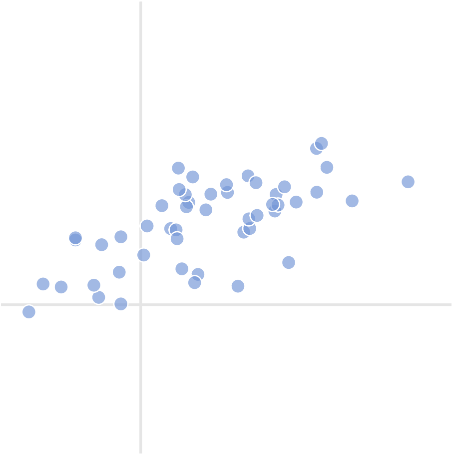
PCA model

Let's talk about how can we give a geometric interpretation of the EVD and change of variable idea

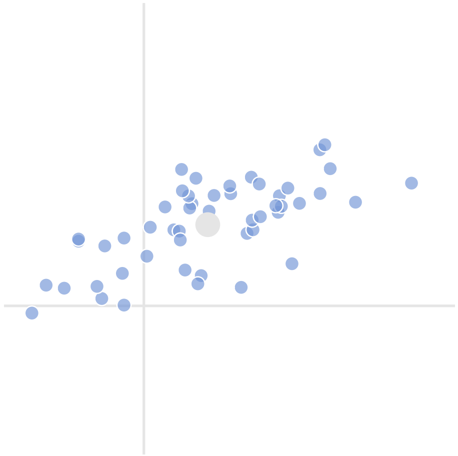
PCA and EVD

The main idea is that the variables in $\mathbf{X}$ are changed into PCs $\mathbf{Z}$. Let's see a toy example for illustration purposes

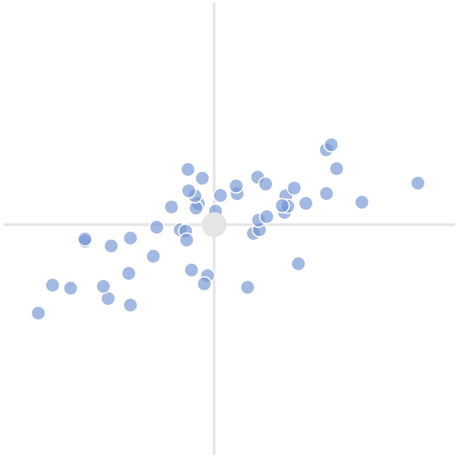
Toy Data (in 2-dimensions)



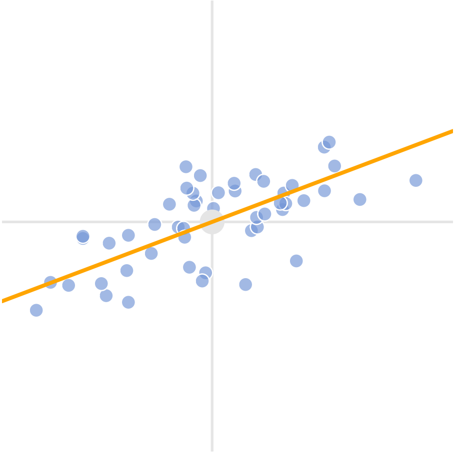
Mean Point (center)



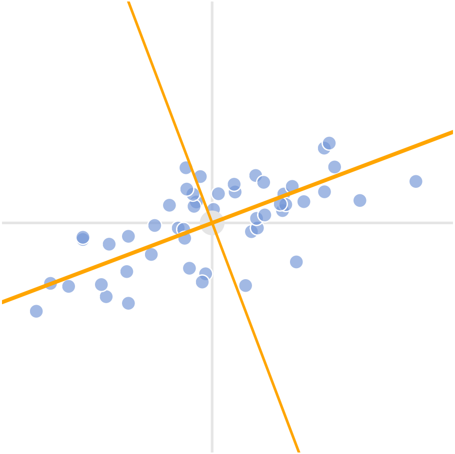
Mean-centering Data



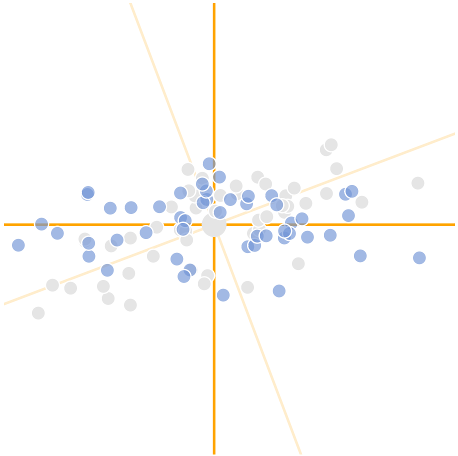
First PC (view as a change of variable)



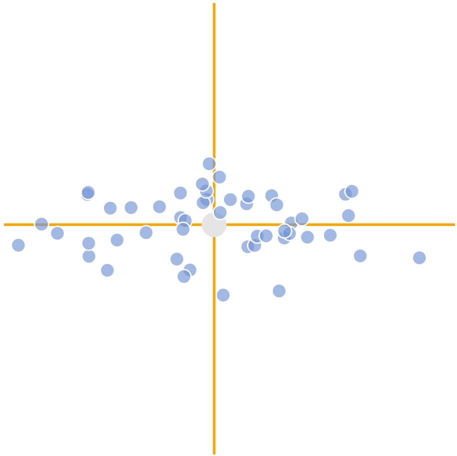
Second PC (view as a change of variable)



Before-and-After Change Comparison



Changed Variables (Rotated Data)



So far we've seen that ...

Recap

Symmetric Matrices

There are several types of quadratic forms, but we are interested in those defined by (real) symmetric matrices like $\mathbf{S} = \mathbf{X}'\mathbf{X}$ (variance-covariance matrix)

Eigenvalue Decomposition

Any symmetric matrix $\mathbf{S}$ can be factorized in terms of its eigenvalue decomposition: $\mathbf{S} = \mathbf{WDW}'$

Principal Components

PCs in matrix format

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

PCs are just linear combinations of $\mathbf{X}$ in which the weights $\mathbf{W}$ are obtained by diagonalizing $\mathbf{S} = \mathbf{X}'\mathbf{X}$

PCA model

We can obtain as many different eigenvalues as the rank of $\mathbf{S}$

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

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

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)