# Principal Components Analysis

The objective in *principal components analysis* (or *PCA*) is to find a new set of variables, with certain desirable properties, that can efficiently represent a set of multiple variables. (Notation follows Manley, with bold = matrix, lower case-italic = scalar, and upper-case italic = implied vector.)

The problem that must be solved in PCA is

given p variables, $X_1, X_2, \ldots, X_p$, find the linear combinations (new variables) $Z_1, Z_2, \ldots, Z_p$, e.g.

$$z_{1i} = a_{11}X_{1i} + a_{12}X_{2i} + \ldots + a_{1p}X_{pi} \qquad i = 1, \ldots, n$$

that have the following properties:

a) $var(Z_1) \geq var(Z_2) \geq \ldots \geq var(Z_p)$

b) maximum simultaneous resemblance of **a** to all $X$'s, and

c) $cov(Z_j Z_k) = 0$, all $k \neq j$

The first, or principal, component

$$z_{1i} = a_{11}X_{1i} + a_{12}X_{2i} + \ldots + a_{1p}X_{pi} \qquad i = 1, \ldots, n$$

is defined by choosing the a's in order to

$$max(\lambda_1) = var(Z_1) = \sum_{i=1}^{n} z_{1i}^2$$

such that $a^2{}_{11} + a^2{}_{12} + \ldots + a^2{}_{1p} = 1$.

A second component

$$z_{2i} = a_{21}X_{1i} + a_{21}X_{2i} + \ldots + a_{2p}X_{pi} \qquad i = 1, \ldots, n$$

can be also be defined to

$$max(\lambda_2) = var(Z_2) = \sum_{i=1}^{n} z_{2i}^2$$

such that $a^2{}_{21} + a^2{}_{21} + \ldots + a^2{}_{2p} = 1$, and $cov(Z_1, Z_2) = 0$.

The optimization problem amounts to requiring the first component to be defined in such a way as to have the maximum *variance* over the $n$ observations. The optimization problem may be stated as:

$$max(\lambda_1) = \frac{1}{n}\sum_{i=1}^{n} z_{1i}^2 \text{ , s.t. } \sum_{j=1}^{p} a_{1j}^2 = 1$$

$$= \frac{1}{n}(\mathbf{a'}_1\mathbf{X})^2 \text{ , s.t. } \mathbf{a'}_1\mathbf{a}_1 = 1$$

$$= \frac{1}{n}(\mathbf{a'}_1\mathbf{X})(\mathbf{X'}\mathbf{a}_1) \text{ , s.t. } \mathbf{a'}_1\mathbf{a}_1 = 1$$

$$= \mathbf{a'}_1\left(\frac{1}{n}\mathbf{XX'}\right)\mathbf{a}_1 \text{ , s.t. } \mathbf{a'}_1\mathbf{a}_1 = 1$$

Note that $\frac{1}{n}\mathbf{XX'} = \mathbf{R}$, the correlation matrix, if

$$(\mu_X = 0) \text{ and } (\sigma_X = 1).$$

Maximizing

$$\lambda_1 = \mathbf{a'}_1\mathbf{R}\mathbf{a}_1 \text{ , s.t. } \mathbf{a'}_1\mathbf{a}_1 = 1$$

is equivalent to maximizing

$$max(u) = \mathbf{a'}_1\mathbf{R}\mathbf{a}_1 - \lambda_1(\mathbf{a'}_1\mathbf{a}_1 - 1),$$

where $\lambda_1$ appears in the equation as a "Lagrange multiplier." To maximize $u$, set the partial derivative of $u$ with respect to **a** to zero:

$$\frac{\partial u}{\partial \mathbf{a}} = 2\mathbf{R}\mathbf{a}_1 - 2\lambda\mathbf{a}_1 = 0 \text{ , or}$$

$$(\mathbf{R} - \lambda_1\mathbf{I})\mathbf{a}_1 = \mathbf{0},$$

This last expression gives $p$ simultaneous equations that must be solved for $\mathbf{a}_1$.

Comparison of this expression with that describing the "eigenstructure" of a square matrix reveals that choosing $\mathbf{a}_1$ to be the first eigenvector of **R**, with $\lambda_1$ as the first eigenvalue solves the optimization problem in PCA, and it also happens that the second and higher components are decribed by the second and higher eigenvectors and eigenvalues of the matrix **R**.

There is a second optimization problem that can be stated that winds up with the same result. In this second problem, instead of choosing the coefficients, $\mathbf{a}_1$, such that the first component has the maximum varibility over the observations, we instead choose $\mathbf{a}_1$ to have the maximum simultaneous resemblance to the X's:

$$\lambda_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{p}(a_{1p}X_{p1})^2 / \sum_{j=1}^{p} a^2{}_{1p}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbf{a'}_1 X^2{}_i / \mathbf{a'}_1\mathbf{a}_1$$

$$= \frac{1}{n}(\mathbf{a'}_1\mathbf{X})^2 / (\mathbf{a'}_1\mathbf{a}_1)$$

$$= \frac{1}{n}(\mathbf{a'}_1\mathbf{XX'}\mathbf{a}_1) / (\mathbf{a'}_1\mathbf{a}_1)$$

$$= \frac{1}{n}(\mathbf{a'}_1\mathbf{R'}\mathbf{a}_1) / (\mathbf{a'}_1\mathbf{a}_1)$$

which leads to the same problem as before:

$$\lambda_1 = \mathbf{a'}_1\mathbf{R}\mathbf{a}_1 \text{ , s.t. } \mathbf{a'}_1\mathbf{a}_1 = 1$$