Title: Canonical Correlation Analysis

Name: Hervé Abdi[1], Vincent Guillemot[2], Aida Eslami[3], Derek Beaton[4]

Affil./Addr. 1: School of Behavioral and Brain Sciences
The University of Texas at Dallas
Richardson, TX 75080, USA
herve@utdallas.edu

Affil./Addr. 2: Institut Pasteur
Bioinformatics and Biostatistics Hub
C3BI, USR 3756 IP CNRS
Paris, France
vincent.guillemot@pasteur.fr

Affil./Addr. 3: Centre for Heart Lung Innovation
University of British Columbia
Vancouver, BC V6Z 1Y6, Canada
aida.eslami@yahoo.fr

Affil./Addr. 4: Rotman Research Institute
Baycrest Health Sciences
Toronto, ON, M6A 2E1 Canada
dbeaton@research.baycrest.org

# Canonical Correlation Analysis

## Synonyms

Canonical analysis. Canonical variate analysis. External factor analysis. [8]

## Glossary

**Canonical correlation** correlation between two canonical variates of the same pair. This is the criterion optimized by CCA.

**Canonical Loadings** Correlation between the original variables and the canonical variates. Sometimes used as a synonym for canonical vectors (because these quantities differ only by their normalization).

**Canonical variates** The latent variables (one per data table) computed in CCA (also called canonical variables, canonical variable scores, or canonical factor scores). The canonical variates have maximal correlation.

**Canonical vectors** The set of coefficients of the linear combinations used to compute the canonical variates, also called canonical *weights*. Canonical vectors are also sometimes called canonical loadings.

**Latent variable** a linear combination of the variables of one data table. In general, a latent variable is computed to satisfy some pre-defined criterion.

## Definition

Canonical correlation analysis (CCA) is a statistical method whose goal is to extract the information common to two data tables that measure quantitative variables on a same set of observations. To do so, CCA creates pairs of linear combinations of the variables (one per table) that have maximal correlation.

## Introduction

Originally defined by Hotelling in 1935 ([17; 18]), canonical correlation analysis (CCA) is a statistical method whose goal is to extract the information common to two data tables that measure quantitative variables on a same set of observations. To do so, CCA computes two sets of linear combinations—called *latent variables*—(one for each data table) that have maximum correlation. To visualize this common information extracted by the analysis, a convenient way is (1) to plot the latent variables of one set against the other set (this creates plots akin to plots of factor scores in principal component analysis); (2) to plot the coefficients of the linear combinations (this creates plots akin to plotting the loadings in principal component analysis); and (3) to plot the correlations between the original variables and the latent variables (this creates "correlation circle" plots like in principal component analysis).

CCA generalizes many standard statistical techniques (e.g., multiple regression, analysis of variance, discriminant analysis) and can also be declined in several related methods that address slightly different types of problems (e.g., different normalization conditions, different types of data).

## Key Points

CCA extracts the information common to two data tables measuring quantitative variables on the same set of observations. For each data table, CCA computes a set of linear combinations of the variables of this table called *latent variables* or *canonical variates* with the constraints that a latent variable from one table has maximal correlation with one latent variable of the other table and no correlation with the remaining latent variables of the other table. The results of the analysis is interpreted using different types of graphical displays that plot the latent variables and the coefficients of the linear combinations used to create the latent variables.

## Historical Background

After principal component analysis (PCA), CCA is one the oldest multivariate techniques that was first defined in 1935 by Hotelling. In addition of being the first method created for the statistical analysis of two data tables, CCA is also of theoretical interest because a very large number of linear multivariate analytic tools are particular cases of CCA. Like most multivariate statistical techniques CCA has been really feasible only with the advent of modern computers. Recent developments involve generalization of CCA to the case of more than two-tables and cross-validation approaches to select important variables and the stability and reliability of the solution obtained on a given sample.

## Canonical Correlation Analysis

## Notations

Matrices are denoted in upper case bold letters, vectors are denoted in lower case bold, and their elements are denoted in lower case italic. Matrices, vectors, and elements from the same matrix all use the same letter (*e.g.,* $\mathbf{A}$, $\mathbf{a}$, $a$). The transpose operation is denoted by the superscript $^{\top}$, the inverse operation is denoted by $^{-1}$. The identity matrix is denoted $\mathbf{I}$, vectors or matrices of ones are denoted $\mathbf{1}$, matrices or vectors of zeros are denoted $\mathbf{0}$. When provided with a square matrix, the diag operator gives a vector with the diagonal elements of this matrix. When provided with a vector, the diag operator gives a diagonal matrix with the elements of the vector as the diagonal elements of this matrix. When provided with a square matrix, the trace operator gives the sum of the diagonal elements of this matrix.

The data tables to be analyzed by CCA of, respectively, size $N \times I$ and $N \times J$, are denoted $\mathbf{X}$ and $\mathbf{Y}$ and collect two sets of, respectively, $I$ and $J$ quantitative measurements obtained on the same $N$ observations. Except if mentioned otherwise, matrices $\mathbf{X}$ and $\mathbf{Y}$ are column centered and normalized, and so:

$$\mathbf{1}^{\top}\mathbf{X} = \mathbf{0},\ \mathbf{1}^{\top}\mathbf{Y} = \mathbf{0}\ , \tag{1}$$

(with $\mathbf{1}$ being a conformable vector of 1s and $\mathbf{0}$ a conformable vector of 0s), and

$$\mathsf{diag}\left\{\mathbf{X}^{\top}\mathbf{X}\right\} = \mathbf{1},\ \mathsf{diag}\left\{\mathbf{Y}^{\top}\mathbf{Y}\right\} = \mathbf{1}\ . \tag{2}$$

Note that because $\mathbf{X}$ and $\mathbf{Y}$ are centered and normalized, matrices, their inner products are correlation matrices that are denoted:

$$\mathbf{R}_{\mathbf{X}} = \mathbf{X}^{\top}\mathbf{X}, \quad \mathbf{R}_{\mathbf{Y}} = \mathbf{Y}^{\top}\mathbf{Y} \text{ and } \mathbf{R} = \mathbf{X}^{\top}\mathbf{Y}\ . \tag{3}$$

## Optimization Problem

In CCA, the problem is to find two latent variables, denoted $\mathbf{f}$ and $\mathbf{g}$ obtained as linear combinations of the columns of, respectively, $\mathbf{X}$ and $\mathbf{Y}$. The coefficients of these linear combinations are stored, respectively, in the $I \times 1$ vector $\mathbf{p}$ and the $J \times 1$ vector $\mathbf{q}$; and, so, we are looking for

$$\mathbf{f} = \mathbf{X}\mathbf{p} \text{ and } \mathbf{g} = \mathbf{Y}\mathbf{q}\ . \tag{4}$$

In CCA, we want the latent variables to have maximum *correlation* and so we are looking for $\mathbf{p}$ and $\mathbf{q}$ satisfying:

$$\delta = \underset{\mathbf{p},\mathbf{q}}{\arg\max}\left\{\operatorname{cor}(\mathbf{f},\mathbf{g})\right\} = \underset{\mathbf{p},\mathbf{q}}{\arg\max}\left\{\frac{\mathbf{f}^{\top}\mathbf{g}}{\sqrt{\left(\mathbf{f}^{\top}\mathbf{f}\right)\left(\mathbf{g}^{\top}\mathbf{g}\right)}}\right\}\ . \tag{5}$$

The correlation between latent variables— maximized by CCA—is called the *canonical correlation* ([13; 18; 16; 20]).

An easy way to solve the maximization problem of CCA is to constraint the latent variables to have unitary norm (because, in this case, the term $\mathbf{f}^{\top}\mathbf{g}$ gives the correlation between $\mathbf{f}$ and $\mathbf{g}$). Specifically, we require that:

$$\mathbf{f}^{\top}\mathbf{f} = \mathbf{p}^{\top}\mathbf{X}^{\top}\mathbf{X}\mathbf{p} = \mathbf{p}^{\top}\mathbf{R}_{\mathbf{X}}\mathbf{p} = 1 = \mathbf{g}^{\top}\mathbf{g} = \mathbf{q}^{\top}\mathbf{Y}^{\top}\mathbf{Y}\mathbf{q} = \mathbf{q}^{\top}\mathbf{R}_{\mathbf{Y}}\mathbf{q}\ . \tag{6}$$

With these notations, the maximization problem from Equation 5 becomes:

$$\arg\max_{\mathbf{p},\mathbf{q}} \left\{ \mathbf{f}^{\top}\mathbf{g} = \mathbf{p}^{\top}\mathbf{R}\mathbf{q} \right\} \text{ under the contraints that } \mathbf{p}^{\top}\mathbf{R}_{\mathbf{X}}\mathbf{p} = 1 = \mathbf{q}^{\top}\mathbf{R}_{\mathbf{Y}}\mathbf{q}\ . \tag{7}$$

### Equivalent Optimum Criteria

The maximization problem expressed in Equation 5 can also be expressed as the following equivalent minimization problem:

$$\arg\min_{\mathbf{p},\mathbf{q}} \left\{ \|\mathbf{X}\mathbf{p} - \mathbf{Y}\mathbf{q}\|_2^2 \right\} = \arg\min_{\mathbf{p},\mathbf{q}} \left\{ \mathsf{trace} \left\{ (\mathbf{X}\mathbf{p} - \mathbf{Y}\mathbf{q})^{\top} (\mathbf{X}\mathbf{p} - \mathbf{Y}\mathbf{q}) \right\} \right\} \tag{8}$$

$$\text{under the contraints that } \mathbf{p}^{\top}\mathbf{R}_{\mathbf{X}}\mathbf{p} = 1 = \mathbf{q}^{\top}\mathbf{R}_{\mathbf{Y}}\mathbf{q}\ .$$

## Lagrangian Approach

A standard way to solve the type of maximization described in Equation 7 is to use a Lagrangian approach. To do so we need: first to define a Lagrangian that incorporates the constraints; next, to take the derivative of the Lagrangian respective to the unknown quantities (here: $\mathbf{p}$ and $\mathbf{q}$); and, finally, to set the derivative to zero in order to obtain the normal equations. Solving the normal equations will then give the values of the parameters that solve the optimization problem (i.e., for CCA: maximum correlation).

Here the Lagrangian (denoted $\mathcal{L}$) that includes the constraints with the two Lagrange multipliers $\alpha$ and $\beta$ is

$$\mathcal{L} = \mathbf{p}^{\top}\mathbf{R}\mathbf{q} - \alpha\left(\mathbf{p}^{\top}\mathbf{R}_{\mathbf{X}}\mathbf{p} - 1\right) - \beta\left(\mathbf{q}^{\top}\mathbf{R}_{\mathbf{Y}}\mathbf{q} - 1\right)\ . \tag{9}$$

The derivatives of the Lagrangian for $\mathbf{p}$ and $\mathbf{q}$ are

$$\frac{\partial\mathcal{L}}{\partial\mathbf{p}} = \mathbf{R}\mathbf{q} - 2\alpha\mathbf{R}_{\mathbf{X}}\mathbf{p} \tag{10}$$

$$\frac{\partial\mathcal{L}}{\partial\mathbf{q}} = \mathbf{R}^{\top}\mathbf{p} - 2\beta\mathbf{R}_{\mathbf{Y}}\mathbf{q}\ . \tag{11}$$

### The Normal Equations

Setting Equations 10 and 11 to zero gives the normal equations:

$$\mathbf{R}\mathbf{q} = 2\alpha\mathbf{R}_{\mathbf{X}}\mathbf{p} \tag{12}$$

$$\mathbf{R}^{\top}\mathbf{p} = 2\beta\mathbf{R}_{\mathbf{Y}}\mathbf{q}\ . \tag{13}$$

### Solution of the Normal Equations

The first step to solve the normal equations is to show that $\alpha = \beta$. This is done by pre-multiplying Equation 12 by $\mathbf{p}$ and Equation 13 by $\mathbf{q}$ to obtain (using the constraints from Equation 3):

$$\mathbf{p}^{\top}\mathbf{R}\mathbf{q} = 2\alpha\mathbf{p}^{\top}\mathbf{R}_{\mathbf{X}}\mathbf{p} \quad = 2\alpha \tag{14}$$

$$\mathbf{q}^{\top}\mathbf{R}^{\top}\mathbf{p} = 2\beta\mathbf{q}^{\top}\mathbf{R}_{\mathbf{Y}}\mathbf{q} = 2\beta\ . \tag{15}$$

Equating $\mathbf{p}^{\top}\mathbf{R}\mathbf{q}$ and $\mathbf{q}^{\top}\mathbf{R}^{\top}\mathbf{p}$ shows that $2\alpha = 2\beta$. For convenience, in what follows we set $\delta = 2\alpha = 2\beta$.

Post-multiplying Equation 12 by $\mathbf{R}_{\mathbf{X}}^{-1}$ and Equation 13 by $\mathbf{R}_{\mathbf{Y}}^{-1}$ gives

$$\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{q} = \delta\mathbf{p} \tag{16}$$

$$\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{p} = \delta\mathbf{q}\ . \tag{17}$$

Replacing $\mathbf{q}$ (respectively $\mathbf{p}$) in Equation 16 (respectively 17) by its expression from 17 (respectively 16) gives the following two eigen-equations (see [3] for a refresher about the eigen-decomposition):

$$\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{p} = \delta^2\mathbf{p} \tag{18}$$

$$\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{q} = \delta^2\mathbf{q}\ , \tag{19}$$

which shows that $\mathbf{p}$ (respectively $\mathbf{q}$) is the eigenvector of the non-symmetric matrix $\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}$ $\left(\text{respectively, } \mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\right)$, associated with the first eigenvalue $\lambda_1 = \delta^2$, and that the maximum correlation (i.e., the *canonical* correlation) is equal to $\delta$. Note that in order to make explicit the constraints expressed in Equation 7, the vectors $\mathbf{p}$ and $\mathbf{q}$ are normalized (respectively) in the metric $\mathbf{R}_{\mathbf{X}}$ and $\mathbf{R}_{\mathbf{Y}}$ (i.e., $\mathbf{p}^{\top}\mathbf{R}_{\mathbf{X}}\mathbf{p} = 1$ and $\mathbf{q}^{\top}\mathbf{R}_{\mathbf{Y}}\mathbf{q} = 1$).

## Additional Pairs of Latent Variables

After the first pair of latent variables has been found, additional pairs of latent variables can be extracted. The criterion from Equations 5 and 7 is still used for the subsequent pairs of latent variables, along with the requirement that the new latent variables are orthogonal to the previous ones. Specifically, if $\mathbf{f}_\ell$ and $\mathbf{g}_\ell$ denote the $\ell$-th pair of latent variables, the orthogonality condition becomes:

$$\mathbf{f}_\ell^{\top}\mathbf{f}_{\ell'} = 0 \quad \text{and} \quad \mathbf{g}_\ell^{\top}\mathbf{g}_{\ell'} = 0 \quad \forall \ell \neq \ell' \; . \tag{20}$$

This orthogonality condition imposed on the latent variables is equivalent to imposing an $\mathbf{R}_{\mathbf{X}}$- (respectively $\mathbf{R}_{\mathbf{Y}}$-) orthogonality condition on the eigenvectors $\mathbf{p}$ and $\mathbf{q}$, namely that:

$$\mathbf{p}_\ell^{\top}\mathbf{R}_{\mathbf{X}}\mathbf{p}_{\ell'} = 0 \quad \text{and} \quad \mathbf{q}_\ell^{\top}\mathbf{R}_{\mathbf{Y}}\mathbf{q}_{\ell'} = 0 \quad \forall \ell \neq \ell' \; . \tag{21}$$

For convenience, latent variables and eigenvectors can be stored in matrices $\mathbf{F}$, $\mathbf{G}$, $\mathbf{P}$, and $\mathbf{Q}$. With these notations, the normalization (from Equation 7) and orthogonality (from Equation 21) conditions are written as

$$\mathbf{F}^{\top}\mathbf{F} = \mathbf{I} \qquad \Longleftrightarrow \qquad \mathbf{P}^{\top}\mathbf{R}_{\mathbf{X}}\mathbf{P} = \mathbf{I} \tag{22}$$

$$\mathbf{G}^{\top}\mathbf{G} = \mathbf{I} \qquad \Longleftrightarrow \qquad \mathbf{Q}^{\top}\mathbf{R}_{\mathbf{Y}}\mathbf{Q} = \mathbf{I} \; . \tag{23}$$

The matrices of eigenvectors $\mathbf{P}$ and $\mathbf{Q}$ are respectively called $\mathbf{R}_{\mathbf{X}}$- and $\mathbf{R}_{\mathbf{Y}}$-orthogonal (the proof of this property is given in the next section, see Equation 27).

The eigen-decompositions for $\mathbf{P}$ and $\mathbf{Q}$ can then be expressed in a matrix form as:

$$\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{P} = \mathbf{P}\boldsymbol{\Lambda} \quad \text{and} \quad \mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{Q} = \mathbf{Q}\boldsymbol{\Lambda}\ . \tag{24}$$

## Solution from the Eigen-decomposition of Symmetric Matrices

The matrices $\mathbf{P}$ and $\mathbf{Q}$ can also be obtained from the decomposition of two symmetric matrices. For example, Matrix $\mathbf{P}$ can be obtained from the following eigen-decomposition

$$\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}} = \widetilde{\mathbf{P}}\boldsymbol{\Lambda}\widetilde{\mathbf{P}}^{\top} \quad \text{with} \quad \widetilde{\mathbf{P}}^{\top}\widetilde{\mathbf{P}} = \mathbf{I}\ . \tag{25}$$

This can be shown by first defining $\widetilde{\mathbf{P}} = \mathbf{R}_{\mathbf{X}}^{\frac{1}{2}}\mathbf{P}$ and replacing $\mathbf{P}$ by $\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{P}}$ in Equation 24 and then simplifying:

$$\begin{aligned}
\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{P} &= \mathbf{P}\boldsymbol{\Lambda} \\
\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{P}} &= \mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{P}}\boldsymbol{\Lambda} \quad \left(\text{because } \mathbf{P} = \mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{P}}\right) \\
\mathbf{R}_{\mathbf{X}}^{\frac{1}{2}}\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{P}} &= \mathbf{R}_{\mathbf{X}}^{\frac{1}{2}}\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{P}}\boldsymbol{\Lambda} \quad \left(\text{Multiply both sides by } \mathbf{R}_{\mathbf{X}}^{\frac{1}{2}}\right) \\
\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{P}} &= \widetilde{\mathbf{P}}\boldsymbol{\Lambda}\ .
\end{aligned} \tag{26}$$

This shows that $\widetilde{\mathbf{P}}$ is the matrix of the eigenvectors of the symmetric matrix $\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}$, which also implies that $\widetilde{\mathbf{P}}^{\top}\widetilde{\mathbf{P}} = \mathbf{I}$. The eigenvectors of the asymmetric matrix $\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}$ are then recovered as $\mathbf{P} = \mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{P}}$. A simple substitution shows that $\mathbf{P}$ is $\mathbf{R}_{\mathbf{X}}$-orthogonal:

$$\mathbf{P}^{\top}\mathbf{R}_{\mathbf{X}}\mathbf{P} = \widetilde{\mathbf{P}}^{\top}\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\mathbf{R}_{\mathbf{X}}\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{P}} = \widetilde{\mathbf{P}}^{\top}\widetilde{\mathbf{P}} = \mathbf{I}. \tag{27}$$

A similar derivation shows that $\mathbf{Q}$ can be obtained from the eigen-decomposition

$$\mathbf{R}_{\mathbf{Y}}^{-\frac{1}{2}}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-\frac{1}{2}} = \widetilde{\mathbf{Q}}\boldsymbol{\Lambda}\widetilde{\mathbf{Q}}^{\top} \quad \text{with} \quad \widetilde{\mathbf{Q}}^{\top}\widetilde{\mathbf{Q}} = \mathbf{I}\ , \tag{28}$$

where $\mathbf{Q} = \mathbf{R}_{\mathbf{Y}}^{-\frac{1}{2}}\widetilde{\mathbf{Q}}$.

## Solution from one Singular Value Decomposition (SVD)

The eigenvector matrices $\mathbf{P}$ and $\mathbf{Q}$ can also be obtained from the singular value decomposition (SVD, see, e.g., [14; 2]) of the matrix

$$\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-\frac{1}{2}} = \widetilde{\mathbf{P}}\boldsymbol{\Delta}\widetilde{\mathbf{Q}}^{\top} = \sum_{\ell}^{L} \delta_{\ell}\widetilde{\mathbf{p}}_{\ell}\widetilde{\mathbf{q}}_{\ell}^{\top} \quad \text{with} \quad \widetilde{\mathbf{P}}^{\top}\widetilde{\mathbf{P}} = \widetilde{\mathbf{Q}}^{\top}\widetilde{\mathbf{Q}} = \mathbf{I}. \tag{29}$$

where $\widetilde{\mathbf{P}}$, $\widetilde{\mathbf{Q}}$, and $\boldsymbol{\Delta}$ denote (respectively) the left, right singular vectors, and a diagonal matrix of the singular values of matrix $\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-\frac{1}{2}}$. The matrices $\mathbf{P}$ and $\mathbf{Q}$ (containing the vectors $\mathbf{p}$ and $\mathbf{q}$) are then computed as

$$\widetilde{\mathbf{P}} = \mathbf{R}_{\mathbf{X}}^{\frac{1}{2}}\mathbf{P} \quad \text{and} \quad \widetilde{\mathbf{Q}} = \mathbf{R}_{\mathbf{Y}}^{\frac{1}{2}}\mathbf{Q} \iff \mathbf{P} = \mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{P}} \quad \text{and} \quad \mathbf{Q} = \mathbf{R}_{\mathbf{Y}}^{-\frac{1}{2}}\widetilde{\mathbf{Q}} \ . \tag{30}$$

### From the eigen-decomposition to the singular value decomposition

To show that $\mathbf{p}$ can be found from the eigen-decompositions from Equations 18 and 19, we first replace $\mathbf{p}$ by $\widetilde{\mathbf{p}}$ in Equation 18. This gives:

$$\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{p}} = \delta^{2}\mathbf{p} \ , \tag{31}$$

then pre-multiplying both sides of Equation 31 by $\mathbf{R}_{\mathbf{X}}^{\frac{1}{2}}$ and simplifying gives

$$\mathbf{R}_{\mathbf{X}}^{\frac{1}{2}}\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{p}} = \delta^{2}\mathbf{R}_{\mathbf{X}}^{\frac{1}{2}}\mathbf{p} \iff \mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\widetilde{\mathbf{p}} = \delta^{2}\widetilde{\mathbf{p}} \ . \tag{32}$$

A similar argument shows that

$$\mathbf{R}_{\mathbf{Y}}^{-\frac{1}{2}}\mathbf{R}^{\top}\mathbf{R}_{\mathbf{X}}^{-1}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-\frac{1}{2}}\widetilde{\mathbf{q}} = \delta^{2}\widetilde{\mathbf{q}} \ . \tag{33}$$

Combining Equations 32 and 33 shows that $\widetilde{\mathbf{p}}$ and $\widetilde{\mathbf{q}}$ are left and right singular vectors of the matrix $\mathbf{R}_{\mathbf{X}}^{-\frac{1}{2}}\mathbf{R}\mathbf{R}_{\mathbf{Y}}^{-\frac{1}{2}}$.

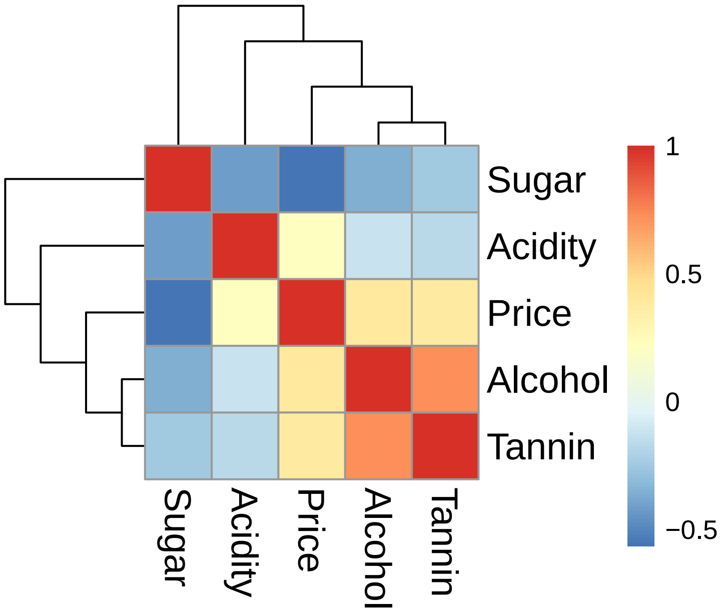


**Fig. 1.** Heatmap of correlation matrix $\mathbf{R_X}$ (i.e., between the variables of matrix $\mathbf{X}$).

## An Example: The Colors and Grapes of Wines

To illustrate CCA we use the data set presented in Table 1. These data describe thirty-six red, rosé, or white wines produced in three different countries (Chili, Canada, and USA) using several different varietal of grapes. These wines are described by two different sets of variables. The first set of variables (i.e., matrix $\mathbf{X}$) describes the objective properties of the wines: Price, Acidity, Alcohol content, Sugar, and Tannin (in what follows we capitalize these descriptors). The second set of variables (i.e., matrix $\mathbf{Y}$) describes the subjective properties of the wines as evaluated by a professional wine taster and consists in ratings on a 9-point rating scale of 8 aspects of taste: fruity, floral, vegetal, spicy, woody, sweet, astringent, acidic, plus an overall evaluation of the hedonic aspect of the wine (i.e., how much the taster liked the wine).

The analysis of this example was performed using the statistical programming language `R` and is available to download from `https://github.com/vguillemot/2Tables`.

Figures 1, 2, and 3 show heatmaps of the correlation matrices $\mathbf{R_X}$, $\mathbf{R_Y}$, and $\mathbf{R}$. As shown in Figure 3 (for matrix $\mathbf{R}$), the objective variables Alcohol and Tannin are

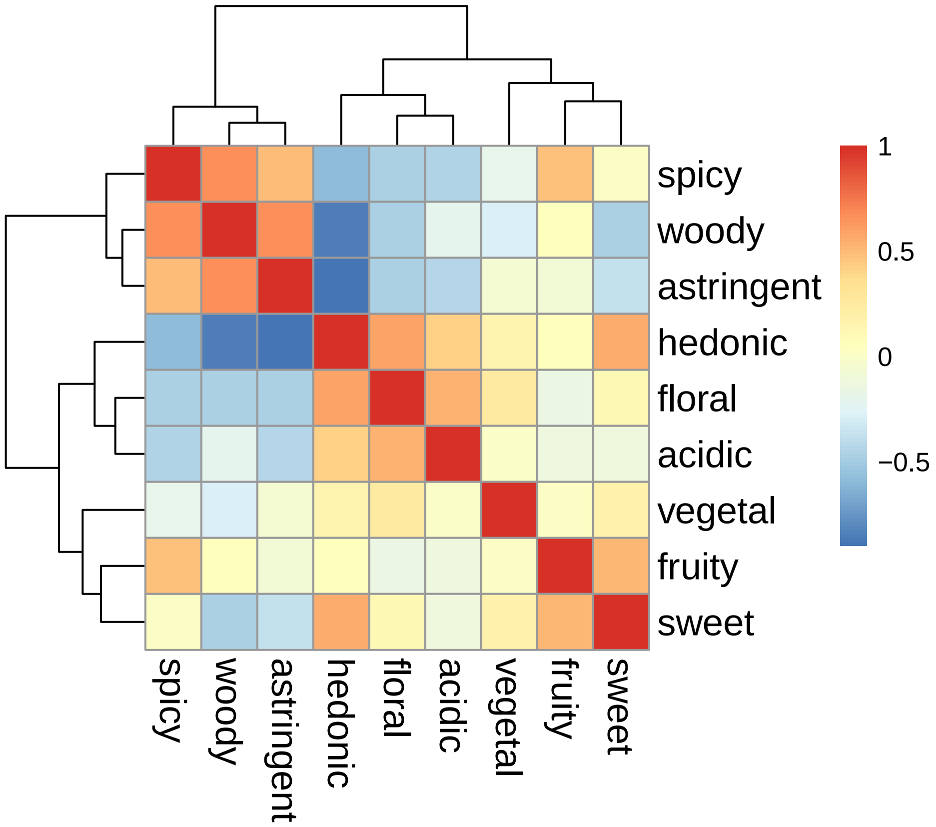


**Fig. 2.** Heatmap of correlation matrix $\mathbf{R_Y}$ (i.e., between the variables of matrix $\mathbf{Y}$).

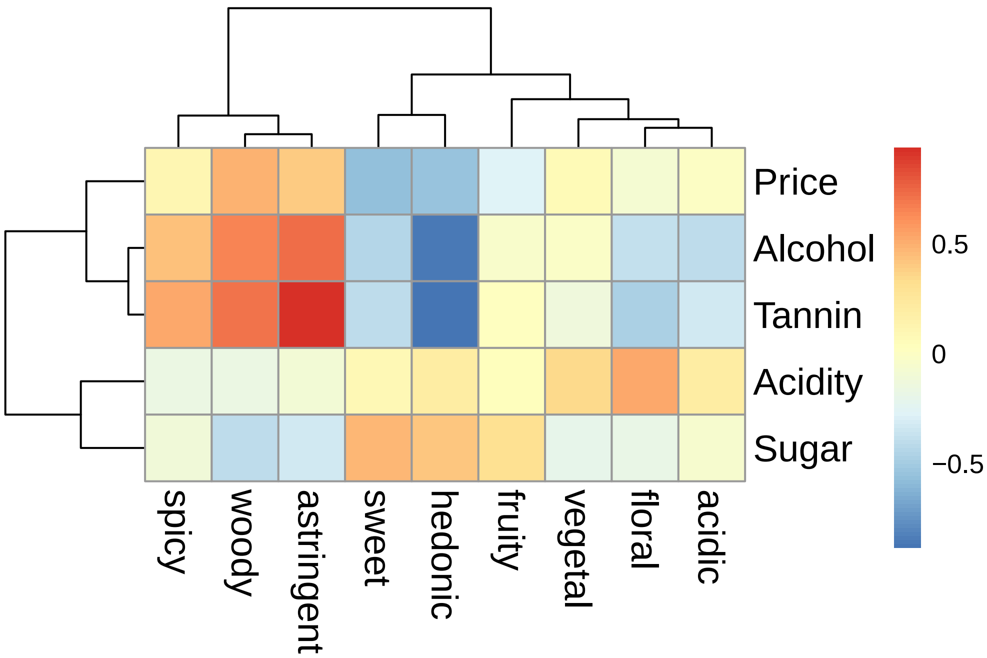


**Fig. 3.** Heatmap of correlation matrix $\mathbf{R}$ (i.e., between the variables of matrices $\mathbf{X}$ and $\mathbf{Y}$).

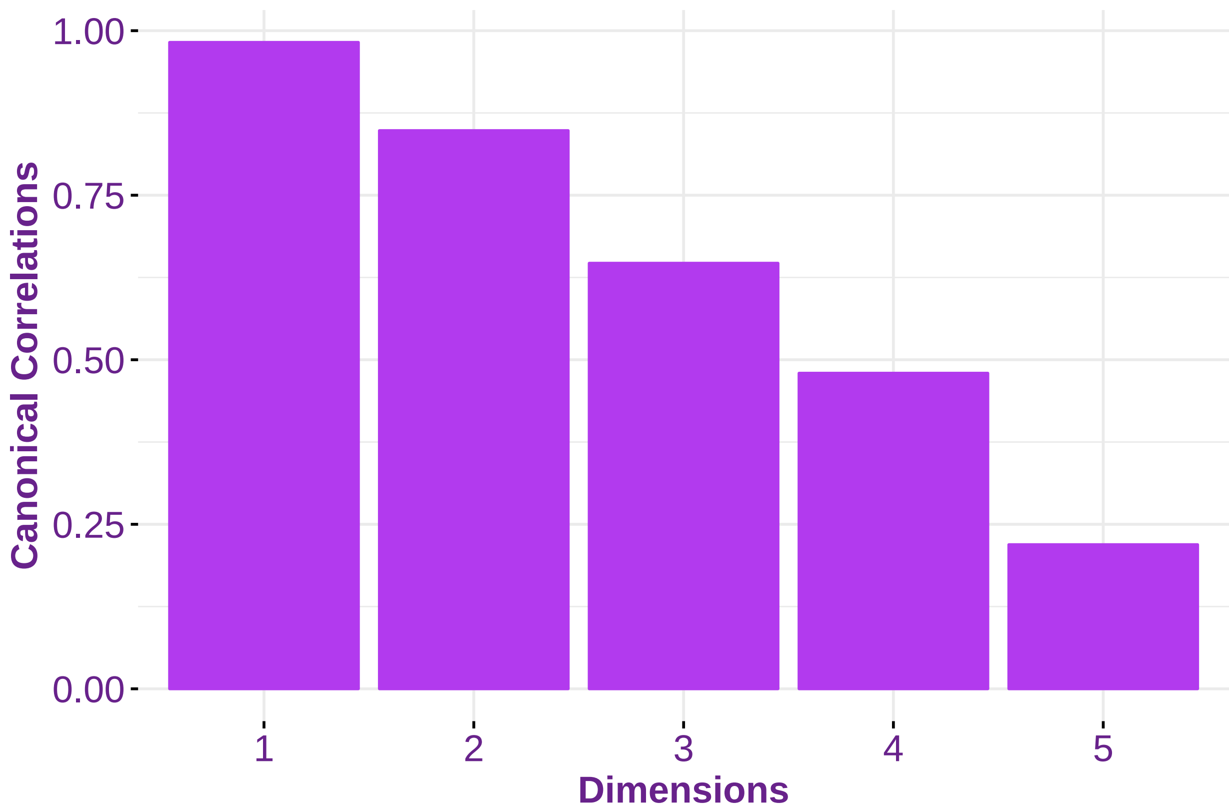


**Fig. 4.** Barplot of the canonical correlations (i.e., correlations between pairs of latent variables for a given dimensions).

positively correlated with the perceived qualities of astringent and woody; by contrast, the perceived hedonic aspect of wine is negatively correlated with Alcohol, Tannin (and Price, so our taster liked inexpensive wines. . . ) and positively correlated with the Sugar content of the wines. Unsurprisingly, the objective amount of Sugar is correlated with the perceived quality sweet.

The CCA of these data found 5 pairs of latent variables [in general CCA will find a maximum of $\min(I, J)$ pairs of latent variables]. The values of the canonical correlations are shown in Figure 4. The first and second canonical correlations are very high (.98 and .85), and so we will only consider them here. As shown in Figures 5 and 6, the latent variables extracted by the analysis are very sensitive to the "the color" of the wines: The first pair of latent variables (Figure 5) isolates the red wines, whereas the second pair of latent variables (Figure 6) roughly orders the wines according to

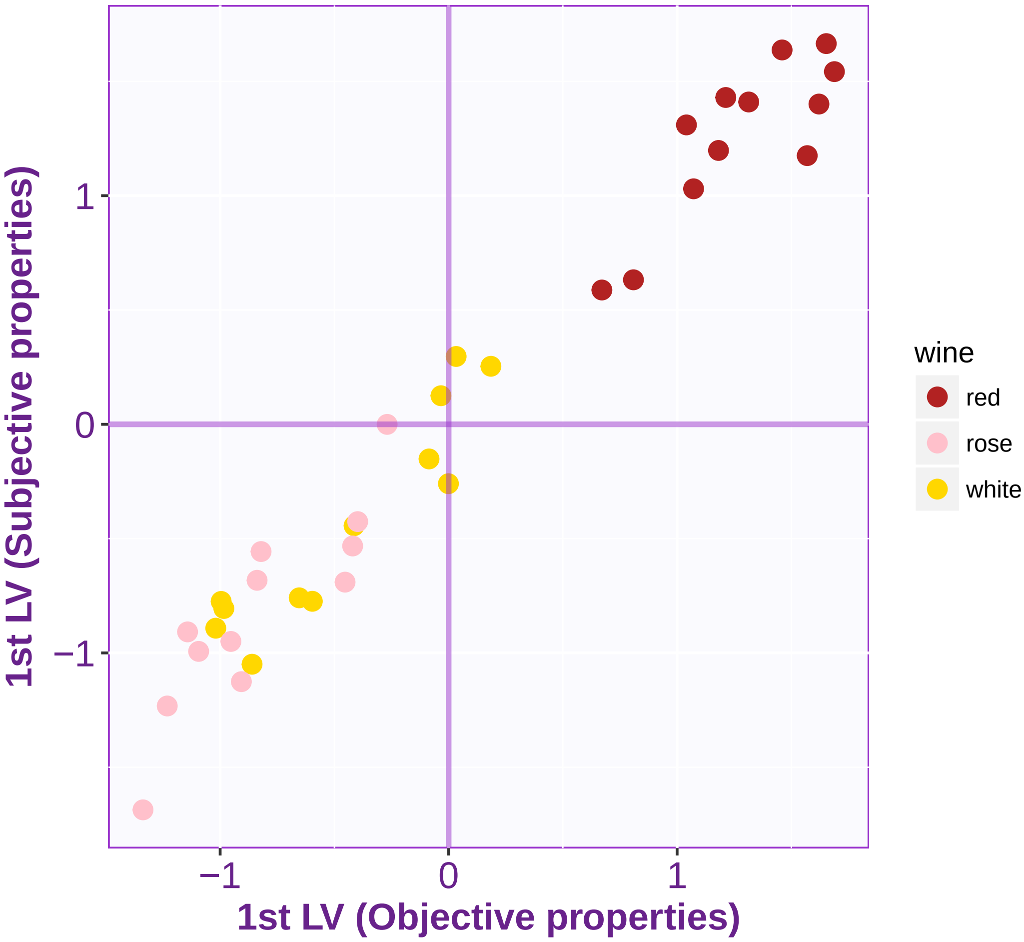


**Fig. 5.** CCA. Latent variables: First latent variable from **X** (1st LV Objective Properties) vs. First latent variable from **Y** (1st LV. Subjective Properties). One point represents one wine. Wines are colored according to their types (i.e., Red, Rosé, or White). Red wines are well separated from the other wines.

their concentration of red pigment (i.e., white, rosé, and red; similar plots using grape varietal or origin of the wines did not show any interesting patterns and are therefore not shown).

To understand the contribution of the variables of **X** and **Y** to the latent variables, two types of plots are used: 1) a plot of the correlation between latent variables and original variables and 2) a plot of the loadings of the variables. Figure 7 (respectively 8) shows the correlation between the original variables (both **X** and **Y**) and respectively **F** (i.e., the latent variables from **X**) and **G** (i.e., the latent variables from **Y**). Figures 9 and 10 display the loadings for, respectively, **X** and **Y** (i.e., matrices

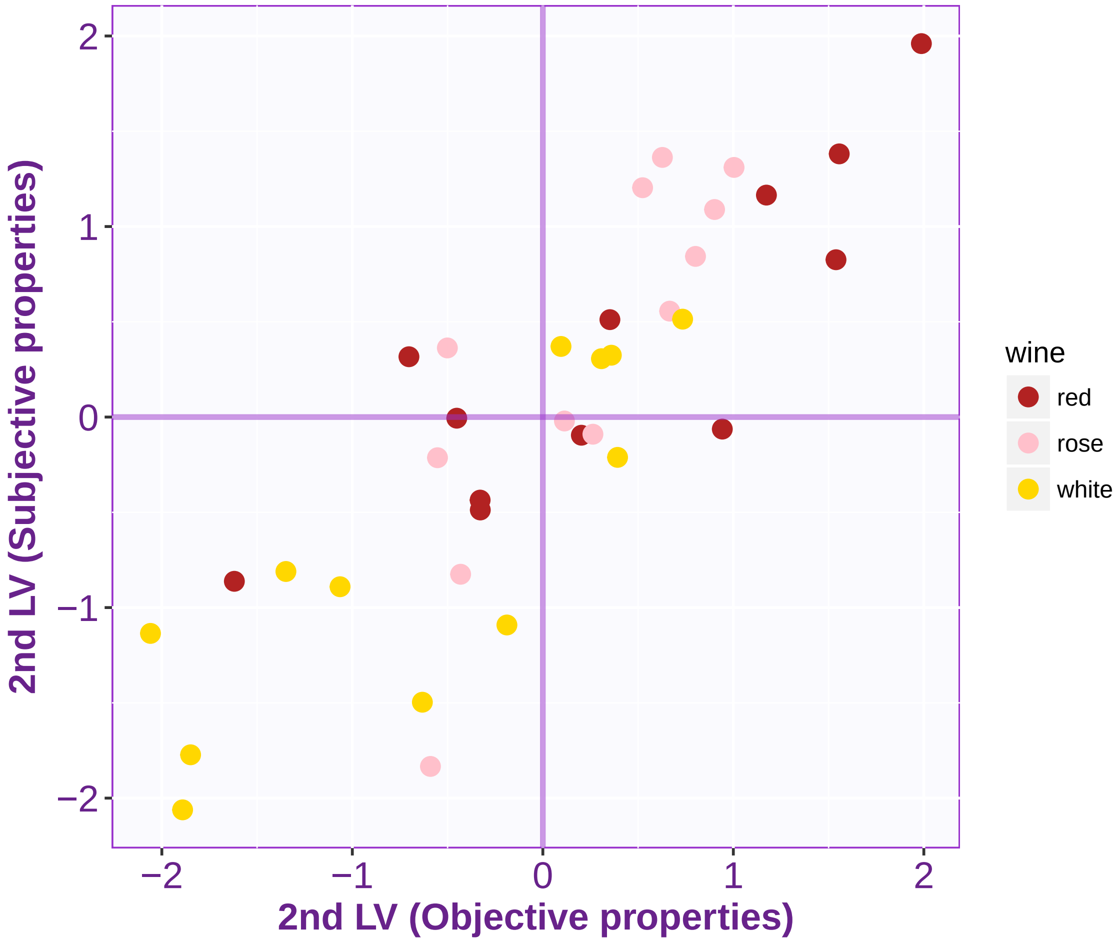


**Fig. 6.** CCA. Latent variables: Second latent variable from **X** (2nd LV Objective Properties) vs. Second latent variable from **Y** (2nd LV. Subjective Properties). One point represents one wine. Wines are colored according to their types (i.e., Red, Rosé, or White).

**P** and **Q**) for the first two dimensions of the analysis. Together these figures indicate that the first dimension reflects the negative correlation between Alcohol (from **X**) and the subjective hedonic evaluation of the wines (from **Y**), whereas the second dimension combines (low) Alcohol, (high) Acidity, and (high) Sugar (from **X**) to reflect their correlations with the subjective variables astringent and hedonic. Figures 7 and 8 show very similar pictures (because the latent variables are very correlated) and this suggests that the first pair of latent variables opposes "bitterness" (i.e., astringent, Alcohol, etc.) to sweetness, whereas the second pair of latent variables opposes bitterness (from astringent) to the "burning" effect of Alcohol.

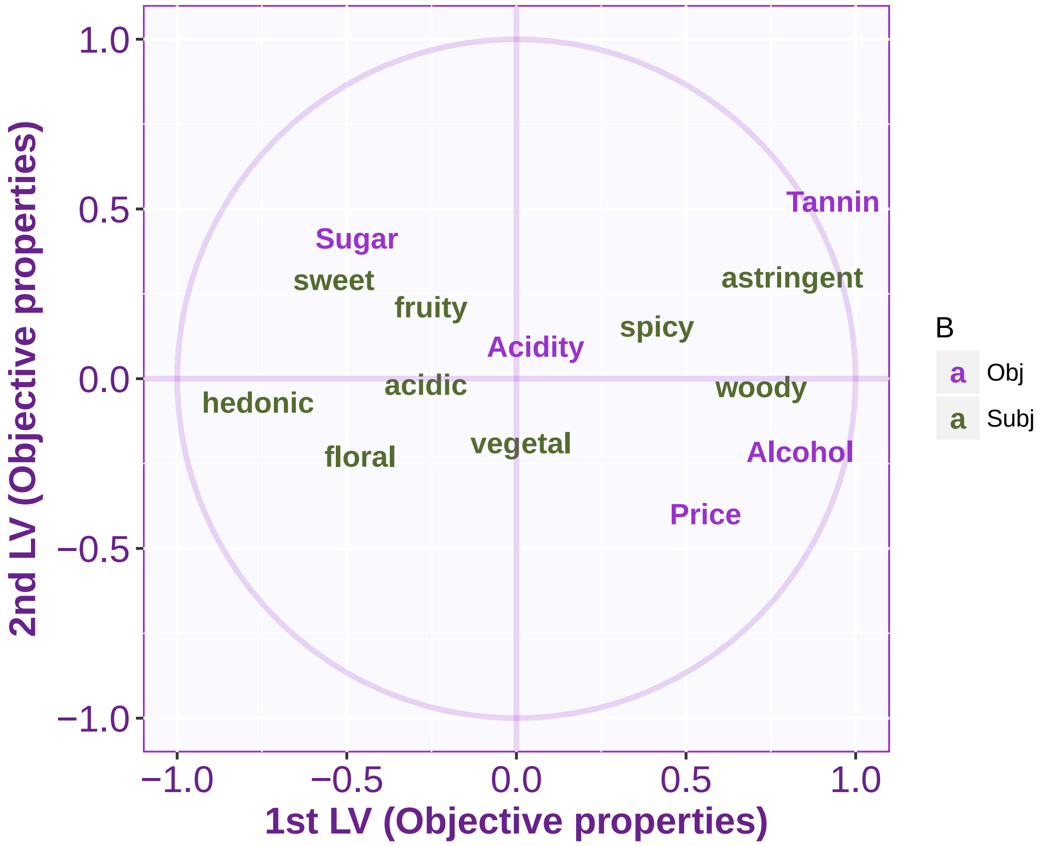


**Fig. 7.** Correlation circle with latent variables from matrix **X**.

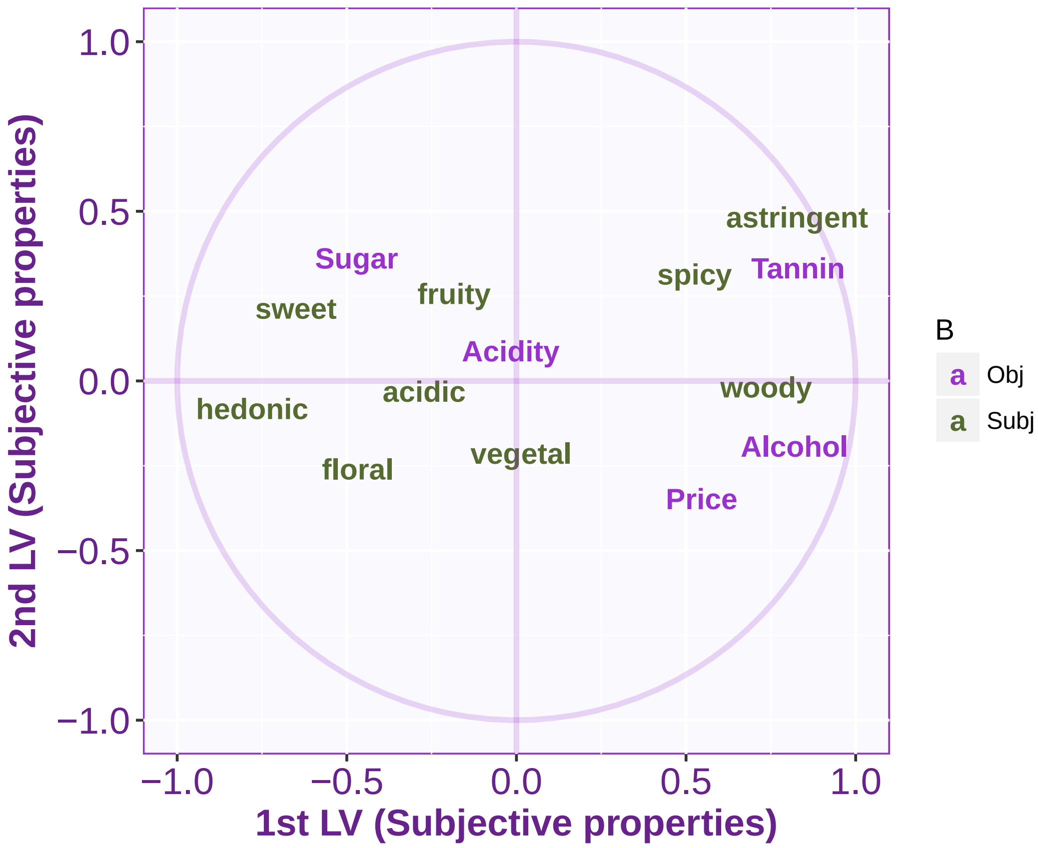


**Fig. 8.** Correlation circle with latent variables from matrix **Y**.

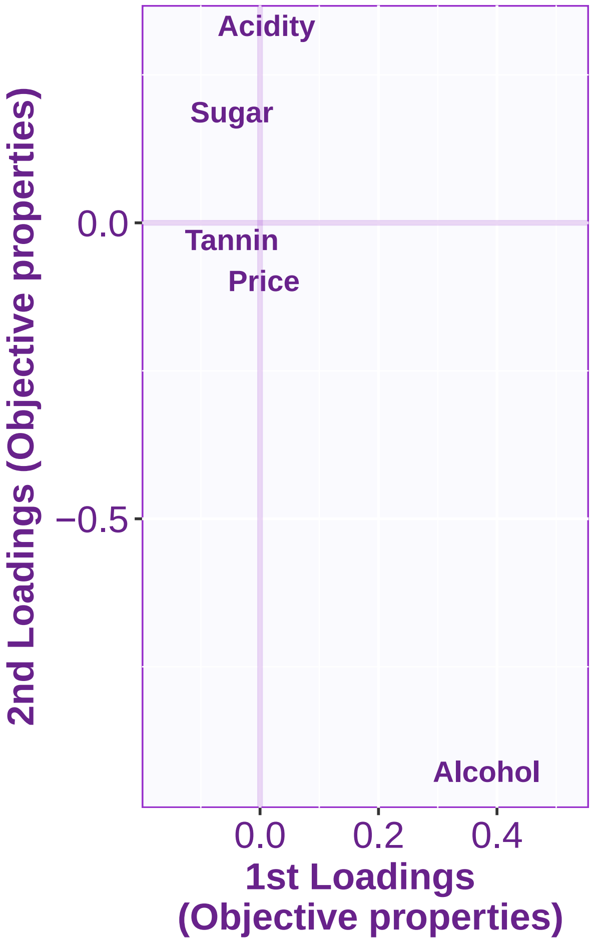


**Fig. 9.** Loadings of the second LV versus the first LV for matrix **X**.

# Variations Over CCA

By imposing slightly different orthogonality conditions than the ones described in Equations 14 and 15, different (but related) alternative methods can be defined to analyze two data tables.

## Inter-Battery Analysis (IBA) et alia

The oldest alternative—originally proposed by Tucker in 1958—called inter-battery analysis (IBA) [27], is also known under a variety of different names such as co-inertia analysis [11], partial least square SVD (PLSVD) [9], partial least square correlation (PLSC) [19; 6], singular value decomposition of the covariance between two fields [10], maximum covariance analysis [29], or even, recently, “multivariate genotype-

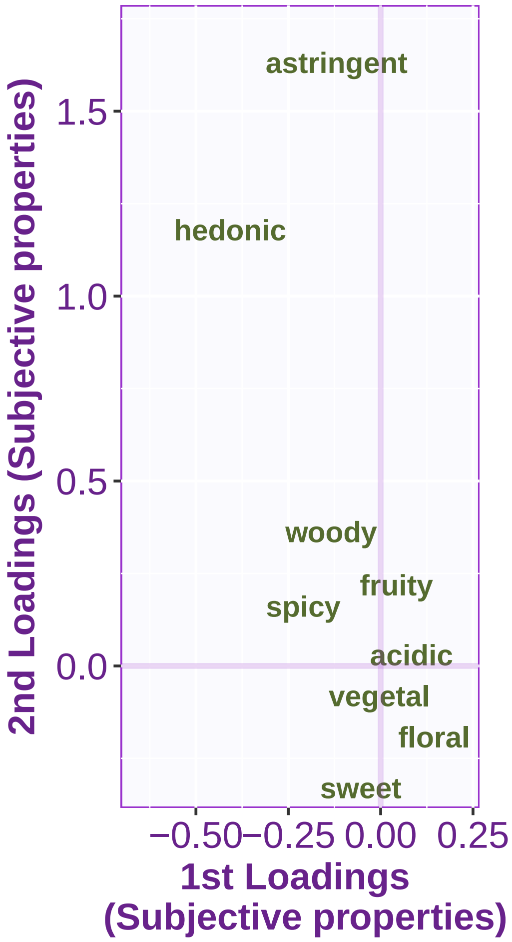


**Fig. 10.** Loadings of the second LV versus the first LV for matrix $\mathbf{Y}$.

phenotype" (MGP) analysis [22]. It is particularly popular in brain-imaging and related domains [21]. In IBA (like in CCA) the latent variables are obtained as linear combinations of the variables of $\mathbf{X}$ and $\mathbf{Y}$ but instead of having maximum correlation (as described in Equations 5 and 7) the latent variables are required to have maximum *covariance*. So we are looking for vectors $\mathbf{p}$ and $\mathbf{q}$ satisfying:

$$\delta = \underset{\mathbf{p},\mathbf{q}}{\arg\max}\left\{\operatorname{cov}\left(\mathbf{f},\mathbf{g}\right)\right\} = \underset{\mathbf{p},\mathbf{q}}{\arg\max}\left\{\mathbf{f}^{\top}\mathbf{g} = \mathbf{p}^{\top}\mathbf{R}\mathbf{q}\ .\right\} \tag{34}$$

An easy way to solve this problem is to impose the constraints that $\mathbf{p}$ and $\mathbf{q}$ have unitary norm, namely that:

$$\mathbf{p}^{\top}\mathbf{p} = 1 = \mathbf{q}^{\top}\mathbf{q}\ . \tag{35}$$

A derivation similar to the one used for CCA shows that the coefficients $\mathbf{P}$ and $\mathbf{Q}$ or the optimal linear combination can be directly obtained from the singular decomposition of the correlation matrix $\mathbf{R}$ as:

$$\mathbf{R} = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{\top} = \sum_{\ell}^{L} \delta_{\ell}\mathbf{p}_{\ell}\mathbf{q}_{\ell}^{\top} \quad \text{with} \quad \mathbf{P}^{\top}\mathbf{P} = \mathbf{Q}^{\top}\mathbf{Q} = \mathbf{I}. \tag{36}$$

where $\mathbf{P}$, $\mathbf{Q}$, and $\boldsymbol{\Delta}$ are (respectively) matrices containing the left, right singular vectors and diagonal matrix of the singular values of $\mathbf{R}$.

Because IBA maximizes the *covariance* between latent variables (instead of the *correlation* for CCA), it does not require the inversion of matrices $\mathbf{R_X}$ and $\mathbf{R_Y}$ and therefore, IBA can be used with rank deficient or badly conditioned matrices and, in particular, when the number of observations is smaller than the number of variables (a configuration sometimes called $N << P$). This makes IBA a robust technique popular for domains using Big or "Wide" Data such as, for example, brain imaging ([19; 15]) or genomics ([7]).

## Asymmetric Two-table Analysis

CCA and IBA are symmetric techniques because the results of the analysis will be unchanged (*mutatis mutandis*) if $\mathbf{X}$ and $\mathbf{Y}$ are permuted. Other related techniques differentiate the roles of $\mathbf{X}$ and $\mathbf{Y}$ and treat $\mathbf{X}$ as a matrix of predictor variables and $\mathbf{Y}$ as a matrix of dependent variables (i.e., to be "explained" or "predicted" by $\mathbf{X}$). Among these techniques, the two most well-known approaches are *redundancy analysis* and *partial least square regression* (see, also, [24] for alternative methods and review).

### Redundancy Analysis (RA)

Redundancy analysis (RA, [28])—originally developed under the name of *principal component analysis of instrumental variables* by Rao in 1964 ([23]) and *simultaneous prediction method* (by Fortier in 1966 [12])—can be seen as a technique intermediate between CCA and IBA. In RA, the vectors stored in $\mathbf{P}$ (i.e., the loadings from $\mathbf{X}$, the predictors) are required to be $\mathbf{R_X}$-normalized (just like in CCA) but the vectors stored in $\mathbf{Q}$ (i.e., the loadings from $\mathbf{Y}$, the predictors) are required to be unit normalized (just like in IBA). Specifically, RA is solving the following optimization problem:

$$\delta = \underset{\mathbf{p},\mathbf{q}}{\arg\max} \left\{ \operatorname{cov}\left(\mathbf{f},\mathbf{g}\right) \right\} = \underset{\mathbf{p},\mathbf{q}}{\arg\max} \left\{ \mathbf{f}^{\top}\mathbf{g} = \mathbf{p}^{\top}\mathbf{R}\mathbf{q} \right\} \quad \text{with} \quad \mathbf{p}^{\top}\mathbf{R}_{\mathbf{X}}\mathbf{p} = 1 = \mathbf{q}^{\top}\mathbf{q}\ . \quad (37)$$

RA can be interpreted as searching for the best predictive linear combinations of the columns of $\mathbf{X}$ or, equivalently, RA is searching for the subspace of $\mathbf{X}$ where the projection of $\mathbf{Y}$ has the largest variance.

### Partial Least Square Regression (PLSR)

Partial Least Square Regression (PLSR)—a technique tailored to cope with multicollinearity of the predictors in a regression framework—(see, e.g., [25; 4] for reviews) starts just like IBA and extracts a first pair of latent variables with maximal covariance, then $\mathbf{f}$ (the latent variable from $\mathbf{X}$) is used, in a regression step, to predict $\mathbf{Y}$. After the first latent variable has been used, its effect is partialled from $\mathbf{X}$ and $\mathbf{Y}$ and the procedure is re-iterated to find subsequent latent variables and loadings.

# Particular cases

## Non Centered CCA: Correspondence analysis

An interesting particular case of *non-centered* CCA concerns the case of group matrices. In a group matrix, the rows represent observations (just like in plain CCA) and the columns represent a set of exclusive groups (i.e., an observation belongs to one and only one group). The group assigned to an observation has a value of 1 for the row representing this observation and all the other columns for this observation (representing the groups not assigned to this observation) will have a value of 0. When both $\mathbf{X}$ and $\mathbf{Y}$ are non-centered and non-normalized group matrices, then the CCA of these matrices will give *correspondence analysis*—a technique developed to analyze contingency tables (see entry on correspondence analysis and, e.g., [14]). When $\mathbf{X}$ and $\mathbf{Y}$ are composed of the concatenation of several non-centered and non-normalized group matrices, the CCA of these two tables will be equivalent to *partial least square correspondence analysis* (PLSCA [7])—a technique originally developed to analyze the information shared by two tables storing qualitative data. In the particular case when $\mathbf{X}$ is composed of the concatenation of several non-centered and non-normalized group matrices, the CCA of $\mathbf{X}$ with itself is equivalent to multiple correspondence analysis.

### Some Other Particular Cases of CCA

CCA is a very general method and so a very large number of methods are particular cases of CCA. For example, when $\mathbf{Y}$ has only one column, CCA becomes (simple and multiple) linear regression. If $\mathbf{X}$ is a group matrix and $\mathbf{Y}$ stores one quantitative variable, CCA becomes analysis of variance. If $\mathbf{Y}$ is a group matrix, CCA becomes discriminant analysis. This versatility of CCA makes it of particular theoretical interest.

## Key Applications

CCA and its derivatives—or variations thereof—are used when the analytic problem is to relate two data tables and this makes these techniques ubiquitous in almost any domains of inquiry from marketing to brain imaging and network analysis (see [5] for examples).

## Future Directions

CCA is still a domain of intense research with future developments likely to be concerned with multi-table extensions (e.g., [16; 26]), “robustification,” and sparsification [30]. All these approaches will make CCA and its related techniques even more suitable for the analysis of very large data sets that are becoming prevalent in analytics.

## Cross-References

Barycentric Discriminant Analysis; Correspondence analysis; Eigenvalues, Singular Value Decomposition; Iterative Methods for Eigenvalues/Eigenvectors; Least Squares; Matrix Algebra, Basics of; Matrix Decomposition; Principal Component Analysis; Regression Analysis; Spectral Analysis;

## References

1. Abdi, H. (2003b). Multivariate analysis. In M. Lewis-Beck, A. Bryman, and T. Futing (Eds), *Encyclopedia for research methods for the social sciences* ()pp. 699–702). Thousand Oaks, CA: Sage.
2. Abdi, H. (2007a). Singular Value Decomposition (SVD) and Generalized Singular Value Decomposition(GSVD). In N.J. Salkind (Ed), *Encyclopedia of measurement and statistics* (pp. 907–912). Thousand Oaks: Sage.
3. Abdi, H. (2007b). Eigen-decomposition: eigenvalues and eigenvectors. In N.J. Salkind (Ed), *Encyclopedia of measurement and statistics* (pp. 304–308). Thousand Oaks: Sage.
4. Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**, 97–106.
5. Abdi, H., Vinzi, V. E., Russolillo, G., Saporta, G., & Trinchera, L. (Eds.). (2016). *The Multiple Facets of Partial Least Squares Methods*. New York: Springer Verlag.
6. Abdi, H., and Williams, L. J. (2013). Partial least squares methods: partial least squares correlation and partial least square regression. *Computational Toxicology*, **II**, 549–579.
7. Beaton, D., Dunlop, J.,ADNI, and Abdi, H. (2016). Partial least squares correspondence analysis: A framework to simultaneously analyze behavioral and genetic data. *Psychological methods*, **21**, 621–651.

8. Bartlett, M.S., (1948). External and internal factor analysis. *British Journal of Psychology*, **1**, 73–81.

9. Bookstein, F. (1994). Partial least squares: a doseresponse model for measurement in the behavioral and brain sciences. *Psycoloquy*, **5(23)**.

10. Bretherton, C. S., Smith, C., and Wallace, J. M. (1992). An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate*, **5**, 541–560.

11. Dolédec, S., and Chessel D. (1994). Co-inertia analysis: An alternative method for studying species-environment relationships. *Freshwater Biology*, **31**, 277–294.

12. Fortier, J.J. (1966). Simultaneous linear prediction. *Psychometrika*, **31**, 369–381.

13. Gittins, R. (2012). *Canonical analysis: a review with applications in ecology.* New York: Springer Verlag.

14. Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* London: Academic Press.

15. Grellmann, C., Bitzer, S., Neumann, J., Westlye, L.T., Andreassen, O.A., Villringer, A.,and Horstmann, A. (2015). Comparison of variants of canonical correlation analysis and partial least squares for combined analysis of MRI and genetic data. *NeuroImage*, **107**, 289–310.

16. Horst, P. (1961). Relations among $m$ sets of measures. *Psychometrika*, **26**, 129–149.

17. Hotteling, H.(1935). The most predicable criterion. *Journal of Educational Psychology*, **26**, 139–142.

18. Hotteling, H.(1936). Relation between two sets of variates. *Biometrika*, **28**, 321–377.

19. Krishnan, A., Williams, L.J., McIntosh, A.R., and Abdi, H. (2010). Partial least squares (PLS) methods for neuroimaging: A tutorial and review. *Neuroimage*, **56**, 455–475.

20. Mardia, K.V., Kent, J.T., and Bibby, J.M. (1980). *Multivariate analysis.* London: Academic Press.

21. McIntosh, A. R., Bookstein, F. L., Haxby, J. V., and Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage*, **3**, 143–157.

22. Mitteroecker, P., Cheverud, J. M., and Pavlicev, M. (2016). Multivariate Analysis of Genotype-Phenotype Association. *Genetics*, genetics.115.181339. http://doi.org/10.1534/genetics.115.181339

23. Rao, C.R. (1964). The use and interpretation of Principal Component Analysis in applied research, *Sankhyā*, **A26**, 329–358.

24. Takane, Y. (2013). *Constrained principal component analysis and related techniques.* Boca Raton: CRC Press.

25. Tenenhaus, M. (1998). *La Régression PLS: Théorie et Pratique.* Paris: Editions Technip.

26. Tenenhaus, A., Philippe, C., Guillemot, V., Le Cao, K.A., Grill, J., & Frouin, V. (2014). Variable selection for generalized canonical correlation analysis. *Biostatistics*,**15**, 569–583.

27. Tucker, L.R. (1958). An inter-battery method of factor analysis. *Psychometrika*, **23**, 111–136.

28. Van Den Wollenberg, A. L. (1977). Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika*, **42**, 207–219.

29. von Storch, H., and Zwiers, F.W. (2002). *Statistical analysis in climate research.* Cambridge (MA): Cambridge University Press.

30. Witten D.M., Tibshirani, R., & Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, **10**, 515–534.

**Table 1.** An example for CCA. Thirty-six wines are described by two sets of variables: objective descriptors (Matrix **X**) and subjective descriptors (Matrix **Y**).

| | | | | Matrix **X**: Objective | | | | | Matrix **Y**: Subjective | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Wine | Origin | Color | Varietal | Price | Acidity | Alcohol | Sugar | Tannin | fruity | floral | vegetal | spicy | woody | Sweet | Astringent | acidic | hedonic |
| CH01 | chili | red | merlot | 11 | 5.33 | 13.80 | 2.75 | 559 | 6 | 2 | 1 | 4 | 5 | 3 | 5 | 4 | 2 |
| CH02 | chili | red | cabernet | 5 | 5.14 | 13.90 | 2.41 | 672 | 5 | 3 | 2 | 3 | 4 | 2 | 6 | 3 | 2 |
| CH03 | chili | red | shiraz | 7 | 5.16 | 14.30 | 2.20 | 455 | 7 | 1 | 2 | 6 | 5 | 3 | 4 | 2 | 2 |
| CH04 | chili | red | pinot | 16 | 4.37 | 13.50 | 3.00 | 348 | 5 | 3 | 2 | 2 | 4 | 1 | 3 | 4 | 4 |
| CH05 | chili | white | chardonnay | 14 | 4.34 | 13.30 | 2.61 | 46 | 5 | 4 | 1 | 3 | 4 | 2 | 1 | 4 | 6 |
| CH06 | chili | white | sauvignon | 8 | 6.60 | 13.30 | 3.17 | 54 | 7 | 5 | 6 | 1 | 1 | 4 | 1 | 5 | 8 |
| CH07 | chili | white | riesling | 9 | 7.70 | 12.30 | 2.15 | 42 | 6 | 7 | 2 | 2 | 2 | 3 | 1 | 6 | 9 |
| CH08 | chili | white | gewurzt | 11 | 6.70 | 12.50 | 2.51 | 51 | 5 | 8 | 2 | 1 | 1 | 4 | 1 | 4 | 9 |
| CH09 | chili | rose | malbec | 4 | 6.50 | 13.00 | 7.24 | 84 | 8 | 4 | 3 | 2 | 2 | 6 | 2 | 3 | 8 |
| CH10 | chili | rose | cabernet | 3 | 4.39 | 12.00 | 4.50 | 90 | 6 | 3 | 2 | 1 | 1 | 5 | 2 | 3 | 8 |
| CH11 | chili | rose | pinot | 6 | 4.89 | 12.00 | 6.37 | 76 | 7 | 2 | 1 | 1 | 1 | 4 | 1 | 4 | 9 |
| CH12 | chili | rose | syrah | 5 | 5.90 | 13.50 | 4.20 | 80 | 8 | 4 | 1 | 3 | 2 | 5 | 2 | 3 | 7 |
| CA01 | canada | red | merlot | 20 | 7.42 | 14.90 | 2.10 | 483 | 5 | 3 | 2 | 3 | 4 | 3 | 4 | 4 | 3 |
| CA02 | canada | red | cabernet | 15 | 7.35 | 14.50 | 1.90 | 698 | 6 | 3 | 2 | 2 | 5 | 2 | 5 | 4 | 2 |
| CA03 | canada | red | shiraz | 20 | 7.50 | 14.50 | 1.50 | 413 | 6 | 2 | 3 | 4 | 3 | 3 | 5 | 1 | 2 |
| CA04 | canada | red | pinot | 25 | 5.70 | 13.30 | 1.70 | 320 | 4 | 2 | 3 | 1 | 3 | 2 | 4 | 4 | 4 |
| CA05 | canada | white | chardonnay | 20 | 6.00 | 13.50 | 3.00 | 35 | 4 | 3 | 2 | 1 | 3 | 2 | 2 | 3 | 5 |
| CA06 | canada | white | sauvignon | 15 | 7.50 | 12.00 | 3.50 | 40 | 8 | 4 | 3 | 2 | 1 | 3 | 1 | 4 | 8 |
| CA07 | canada | white | riesling | 15 | 7.00 | 11.90 | 3.40 | 48 | 7 | 5 | 1 | 1 | 3 | 3 | 1 | 7 | 8 |
| CA08 | canada | white | gewurzt | 18 | 6.30 | 13.90 | 2.80 | 39 | 6 | 5 | 2 | 2 | 2 | 3 | 2 | 5 | 6 |
| CA09 | canada | rose | malbec | 8 | 5.90 | 12.00 | 5.50 | 90 | 6 | 3 | 3 | 3 | 2 | 4 | 2 | 4 | 8 |
| CA10 | canada | rose | cabernet | 6 | 5.60 | 12.50 | 4.00 | 85 | 5 | 4 | 1 | 3 | 2 | 4 | 2 | 4 | 7 |
| CA11 | canada | rose | pinot | 9 | 6.20 | 13.00 | 6.00 | 75 | 5 | 3 | 2 | 1 | 2 | 3 | 2 | 3 | 7 |
| CA12 | canada | rose | syrah | 9 | 5.80 | 13.00 | 3.50 | 83 | 7 | 3 | 2 | 3 | 3 | 4 | 1 | 4 | 7 |
| US01 | usa | red | merlot | 25 | 6.00 | 13.60 | 3.50 | 578 | 7 | 2 | 2 | 5 | 6 | 3 | 4 | 3 | 2 |
| US02 | usa | red | cabernet | 15 | 6.50 | 14.60 | 3.50 | 710 | 8 | 3 | 1 | 4 | 5 | 3 | 5 | 3 | 2 |
| US03 | usa | red | shiraz | 25 | 5.30 | 13.90 | 1.99 | 610 | 8 | 2 | 3 | 7 | 6 | 4 | 5 | 3 | 1 |
| US04 | usa | red | pinot | 28 | 6.10 | 14.00 | 0.00 | 340 | 6 | 3 | 2 | 2 | 5 | 2 | 4 | 4 | 2 |
| US05 | usa | white | chardonnay | 15 | 7.20 | 13.30 | 1.10 | 41 | 6 | 4 | 2 | 3 | 6 | 3 | 2 | 4 | 5 |
| US06 | usa | white | sauvignon | 8 | 7.20 | 13.50 | 1.00 | 50 | 6 | 5 | 5 | 1 | 2 | 4 | 2 | 4 | 7 |
| US07 | usa | white | riesling | 10 | 8.60 | 12.00 | 1.65 | 47 | 5 | 5 | 3 | 2 | 2 | 4 | 2 | 5 | 8 |
| US08 | usa | white | gewurzt | 20 | 9.60 | 12.00 | 0.00 | 45 | 6 | 6 | 3 | 2 | 2 | 4 | 2 | 3 | 8 |
| US09 | usa | rose | malbec | 3 | 6.20 | 12.50 | 4.00 | 84 | 8 | 2 | 1 | 4 | 3 | 5 | 2 | 4 | 7 |
| US10 | usa | rose | cabernet | 4 | 5.71 | 12.50 | 4.30 | 93 | 8 | 3 | 3 | 3 | 2 | 6 | 2 | 3 | 8 |
| US11 | usa | rose | pinot | 8 | 5.40 | 13.00 | 3.10 | 79 | 6 | 1 | 1 | 2 | 3 | 4 | 1 | 3 | 6 |
| US12 | usa | rose | syrah | 6 | 6.50 | 13.50 | 3.00 | 89 | 9 | 3 | 2 | 5 | 4 | 3 | 2 | 3 | 5 |

**Table 2.** An example for CCA. Canonical correlations ($\delta_\ell$), Loadings for matrices $\mathbf{X}$ (Objective descriptors, loading matrix $\mathbf{P}$) and $\mathbf{Y}$ (Subjective descriptors loading matrix $\mathbf{Q}$) for the 5 dimensions extracted by CCA.

| | | Matrix $\mathbf{P}$: Objective | | | | | Matrix $\mathbf{Q}$: Subjective | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dimension | $\delta_\ell$ | Price | Acidity | Alcohol | Sugar | Tannin | fruity | floral | vegetal | spicy | woody | sweet | astringent | acidic | hedonic |
| 1 | .98 | 0.026 | -0.088 | 0.489 | -0.134 | 0.002 | 0.000 | 0.150 | 0.030 | -0.068 | -0.109 | 0.082 | -0.095 | 0.059 | -0.586 |
| 2 | .85 | -0.062 | 0.306 | -0.927 | 0.188 | 0.006 | 0.274 | -0.247 | -0.150 | 0.114 | 0.307 | -0.285 | 1.688 | 0.085 | 1.179 |
| 3 | .65 | -0.024 | 0.678 | -0.243 | -0.288 | 0.001 | 0.044 | 0.605 | 0.366 | 0.024 | 0.390 | -0.133 | 0.465 | -0.083 | 0.327 |
| 4 | .48 | -0.057 | 0.574 | 1.382 | 0.428 | -0.002 | 0.565 | 0.368 | -0.076 | -0.427 | -0.505 | 0.584 | -0.531 | -0.423 | -0.870 |
| 5 | .22 | 0.157 | 0.360 | -0.131 | 0.627 | -0.000 | 0.744 | -0.205 | 0.087 | -0.877 | 0.290 | -0.568 | 0.296 | 0.065 | 0.185 |