Chapter 16
Canonical Correlation Analysis

Complex multivariate data structures are better understood by studying low-dimensional projections. For a joint study of two data sets, we may ask what type of low-dimensional projection helps in finding possible joint structures for the two samples. The canonical correlation analysis (CCA) is a standard tool of multivariate statistical analysis for discovery and quantification of associations between two sets of variables.

The basic technique is based on projections. One defines an index (projected multivariate variable) that maximally correlates with the index of the other variable for each sample separately. The aim of CCA is to maximise the association (measured by correlation) between the low-dimensional projections of the two data sets. The canonical correlation vectors are found by a joint covariance analysis of the two variables. The technique is applied to a marketing example where the association of a price factor and other variables (like design, sportiness etc.) is analysed. Tests are given on how to evaluate the significance of the discovered association.

16.1 Most Interesting Linear Combination

The associations between two sets of variables may be identified and quantified by CCA. The technique was originally developed by Hotelling (1935) who analysed how arithmetic speed and arithmetic power are related to reading speed and reading power. Other examples are the relation between governmental policy variables and economic performance variables and the relation between job and company characteristics.

W.K. Härdle, L. Simar, *Applied Multivariate Statistical Analysis*,
DOI 10.1007/978-3-662-45171-7_16

Suppose we are given two random variables $X \in \mathbb{R}^q$ and $Y \in \mathbb{R}^p$. The idea is to find an index describing a (possible) link between X and Y. CCA is based on linear indices, i.e. linear combinations

$$a^\top X \quad \text{and} \quad b^\top Y$$

of the random variables. CCA searches for vectors a and b such that the relation of the two indices $a^\top x$ and $b^\top y$ is quantified in some interpretable way. More precisely, one is looking for the "most interesting" projections a and b in the sense that they maximise the correlation

$$\rho(a, b) = \rho_{a^\top X\, b^\top Y} \tag{16.1}$$

between the two indices.

Let us consider the correlation $\rho(a, b)$ between the two projections in more detail. Suppose that

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \left(\begin{pmatrix} \mu \\ \nu \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \right)$$

where the sub-matrices of this covariance structure are given by

$$\begin{aligned} \mathsf{Var}(X) &= \Sigma_{XX}\ (q \times q) \\ \mathsf{Var}(Y) &= \Sigma_{YY}\ (p \times p) \\ \mathsf{Cov}(X, Y) &= \mathsf{E}(X - \mu)(Y - \nu)^\top = \Sigma_{XY} = \Sigma_{YX}^\top \quad (q \times p). \end{aligned}$$

Using (3.7) and (4.26),

$$\rho(a, b) = \frac{a^\top \Sigma_{XY} b}{(a^\top \Sigma_{XX} a)^{1/2}\ (b^\top \Sigma_{YY} b)^{1/2}}\,. \tag{16.2}$$

Therefore, $\rho(ca, b) = \rho(a, b)$ for any $c \in \mathbb{R}^+$. Given the invariance of scale we may rescale projections a and b and thus we can equally solve

$$\max_{a,b} \quad a^\top \Sigma_{XY} b$$

under the constraints

$$\begin{aligned} a^\top \Sigma_{XX} a &= 1 \\ b^\top \Sigma_{YY} b &= 1. \end{aligned}$$

For this problem, define

$$\mathcal{K} = \Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1/2}. \tag{16.3}$$

Recall the singular value decomposition of $\mathcal{K}(q \times p)$ from Theorem 2.2. The matrix $\mathcal{K}$ may be decomposed as

$$\mathcal{K} = \Gamma \Lambda \Delta^\top$$

with

$$\begin{aligned} \Gamma &= (\gamma_1, \ldots, \gamma_k) \\ \Delta &= (\delta_1, \ldots, \delta_k) \\ \Lambda &= \operatorname{diag}(\lambda_1^{1/2}, \ldots, \lambda_k^{1/2}) \end{aligned} \tag{16.4}$$

where by (16.3) and (2.15),

$$k = \operatorname{rank}(\mathcal{K}) = \operatorname{rank}(\Sigma_{XY}) = \operatorname{rank}(\Sigma_{YX}) \,,$$

and $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_k$ are the nonzero eigenvalues of $\mathcal{N}_1 = \mathcal{K}\mathcal{K}^\top$ and $\mathcal{N}_2 = \mathcal{K}^\top \mathcal{K}$ and γ_i and δ_j are the standardised eigenvectors of $\mathcal{N}_1$ and $\mathcal{N}_2$ respectively.

Define now for $i = 1, \ldots, k$ the vectors

$$a_i = \Sigma_{XX}^{-1/2} \gamma_i, \tag{16.5}$$

$$b_i = \Sigma_{YY}^{-1/2} \delta_i, \tag{16.6}$$

which are called the *canonical correlation vectors*. Using these canonical correlation vectors we define the *canonical correlation variables*

$$\eta_i = a_i^\top X \tag{16.7}$$

$$\varphi_i = b_i^\top Y. \tag{16.8}$$

The quantities $\rho_i = \lambda_i^{1/2}$ for $i = 1, \ldots, k$ are called the *canonical correlation coefficients*.

From the properties of the singular value decomposition given in (16.4) we have

$$\mathsf{Cov}(\eta_i, \eta_j) = a_i^\top \Sigma_{XX} a_j = \gamma_i^\top \gamma_j = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases} \tag{16.9}$$

The same is true for $\mathsf{Cov}(\varphi_i, \varphi_j)$. The following theorem tells us that the canonical correlation vectors are the solution to the maximisation problem of (16.1).

Theorem 16.1 *For any given r, $1 \leq r \leq k$, the maximum*

$$C(r) = \max_{a,b} a^\top \Sigma_{XY} b \tag{16.10}$$

subject to

$$a^\top \Sigma_{XX} a = 1, \quad b^\top \Sigma_{YY} b = 1$$

and

$$a_i^\top \Sigma_{XX} a = 0 \text{ for } i = 1, \ldots, r-1$$

is given by

$$C(r) = \rho_r = \lambda_r^{1/2}$$

and is attained when $a = a_r$ and $b = b_r$.

Proof The proof is given in three steps.

(i) Fix a and maximise over b, i.e. solve:

$$\max_b \left(a^\top \Sigma_{XY} b\right)^2 = \max_b \left(b^\top \Sigma_{YX} a\right)\left(a^\top \Sigma_{XY} b\right)$$

subject to $b^\top \Sigma_{YY} b = 1$. By Theorem 2.5 the maximum is given by the largest eigenvalue of the matrix

$$\Sigma_{YY}^{-1} \Sigma_{YX} a a^\top \Sigma_{XY}.$$

By Corollary 2.2, the only nonzero eigenvalue equals

$$a^\top \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} a. \tag{16.11}$$

(ii) Maximise (16.11) over a subject to the constraints of the theorem. Put $\gamma = \Sigma_{XX}^{1/2} a$ and observe that (16.11) equals

$$\gamma^\top \Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-1/2} \gamma = \gamma^\top \mathcal{K}^\top \mathcal{K} \gamma.$$

Thus, solve the equivalent problem

$$\max_\gamma \gamma^\top \mathcal{N}_1 \gamma \tag{16.12}$$

subject to $\gamma^\top \gamma = 1$, $\gamma_i^\top \gamma = 0$ for $i = 1, \ldots, r-1$.

Note that the γ_i's are the eigenvectors of $\mathcal{N}_1$ corresponding to its first $r-1$ largest eigenvalues. Thus, as in Theorem 11.3, the maximum in (16.12) is obtained by setting γ equal to the eigenvector corresponding to the rth largest eigenvalue, i.e. $\gamma = \gamma_r$ or equivalently $a = a_r$. This yields

$$C^2(r) = \gamma_r^\top \mathcal{N}_1 \gamma_r = \lambda_r \gamma_r^\top \gamma = \lambda_r.$$

(iii) Show that the maximum is attained for $a = a_r$ and $b = b_r$. From the SVD of $\mathcal{K}$ we conclude that $\mathcal{K}\delta_r = \rho_r \gamma_r$ and hence

$$a_r^\top \Sigma_{XY} b_r = \gamma_r^\top \mathcal{K} \delta_r = \rho_r \gamma_r^\top \gamma_r = \rho_r.$$

□

Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \left(\begin{pmatrix} \mu \\ \nu \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \right).$$

The canonical correlation vectors

$$a_1 = \Sigma_{XX}^{-1/2} \gamma_1,$$
$$b_1 = \Sigma_{YY}^{-1/2} \delta_1$$

maximise the correlation between the canonical variables

$$\eta_1 = a_1^\top X,$$
$$\varphi_1 = b_1^\top Y.$$

The covariance of the canonical variables η and φ is given in the next theorem.

Theorem 16.2 *Let η_i and φ_i be the ith canonical correlation variables ($i = 1, \ldots, k$). Define $\eta = (\eta_1, \ldots, \eta_k)$ and $\varphi = (\varphi_1, \ldots, \varphi_k)$. Then*

$$\mathsf{Var} \begin{pmatrix} \eta \\ \varphi \end{pmatrix} = \begin{pmatrix} \mathcal{I}_k & \Lambda \\ \Lambda & \mathcal{I}_k \end{pmatrix}$$

with Λ given in (16.4).

This theorem shows that the canonical correlation coefficients, $\rho_i = \lambda_i^{1/2}$, are the covariances between the canonical variables η_i and φ_i and that the indices $\eta_1 = a_1^\top X$ and $\varphi_1 = b_1^\top Y$ have the maximum covariance $\sqrt{\lambda_1} = \rho_1$.

The following theorem shows that canonical correlations are invariant w.r.t. linear transformations of the original variables.

Theorem 16.3 *Let $X^* = \mathcal{U}^\top X + u$ and $Y^* = \mathcal{V}^\top Y + v$ where $\mathcal{U}$ and $\mathcal{V}$ are nonsingular matrices. Then the canonical correlations between X^* and Y^* are the same as those between X and Y. The canonical correlation vectors of X^* and Y^* are given by*

$$\begin{aligned} a_i^* &= \mathcal{U}^{-1} a_i, \\ b_i^* &= \mathcal{V}^{-1} b_i. \end{aligned} \tag{16.13}$$

Summary

- ↪ CCA aims to identify possible links between two (sub-)sets of variables $X \in \mathbb{R}^q$ and $Y \in \mathbb{R}^p$. The idea is to find indices $a^\top X$ and $b^\top Y$ such that the correlation $\rho(a, b) = \rho_{a^\top X b^\top Y}$ is maximal.
- ↪ The maximum correlation (under constraints) is attained by setting $a_i = \Sigma_{XX}^{-1/2} \gamma_i$ and $b_i = \Sigma_{YY}^{-1/2} \delta_i$, where γ_i and δ_i denote the eigenvectors of $\mathcal{K}\mathcal{K}^\top$ and $\mathcal{K}^\top \mathcal{K}$, $\mathcal{K} = \Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1/2}$ respectively.
- ↪ The vectors a_i and b_i are called canonical correlation vectors.
- ↪ The indices $\eta_i = a_i^\top X$ and $\varphi_i = b_i^\top Y$ are called canonical correlation variables.
- ↪ The values $\rho_1 = \sqrt{\lambda_1}, \ldots, \rho_k = \sqrt{\lambda_k}$, which are the square roots of the nonzero eigenvalues of $\mathcal{K}\mathcal{K}^\top$ and $\mathcal{K}^\top \mathcal{K}$, are called the canonical correlation coefficients. The covariance between the canonical correlation variables is $\mathsf{Cov}(\eta_i, \varphi_i) = \sqrt{\lambda_i}$, $i = 1, \ldots, k$.
- ↪ The first canonical variables, $\eta_1 = a_1^\top X$ and $\varphi_1 = b_1^\top Y$, have the maximum covariance $\sqrt{\lambda_1}$.
- ↪ Canonical correlations are invariant w.r.t. linear transformations of the original variables X and Y.

16.2 Canonical Correlation in Practice

In practice we have to estimate the covariance matrices Σ_{XX}, Σ_{XY} and Σ_{YY}. Let us apply the CCA to the car marks data (see Table 22.7). In the context of this data set one is interested in relating price variables with variables such as sportiness and safety. In particular, we would like to investigate the relation between the two variables *non-depreciation of value* and *price of the car* and all other variables.

Example 16.1 We perform the CCA on the data matrices $\mathcal{X}$ and $\mathcal{Y}$ that correspond to the set of values {Price, Value Stability} and {Economy, Service, Design, Sporty car, Safety, Easy handling}, respectively. The estimated covariance matrix $\mathcal{S}$ is given by

$$\mathcal{S} = \begin{array}{c} \begin{array}{cc|cccccc} \text{Price} & \text{Value} & \text{Econ.} & \text{Serv.} & \text{Design} & \text{Sport.} & \text{Safety} & \text{Easy h.} \end{array} \\ \left(\begin{array}{cc|cccccc} 1.41 & -1.11 & 0.78 & -0.71 & -0.90 & -1.04 & -0.95 & 0.18 \\ -1.11 & 1.19 & -0.42 & 0.82 & 0.77 & 0.90 & 1.12 & 0.11 \\ \hline 0.78 & -0.42 & 0.75 & -0.23 & -0.45 & -0.42 & -0.28 & 0.28 \\ -0.71 & 0.82 & -0.23 & 0.66 & 0.52 & 0.57 & 0.85 & 0.14 \\ -0.90 & 0.77 & -0.45 & 0.52 & 0.72 & 0.77 & 0.68 & -0.10 \\ -1.04 & 0.90 & -0.42 & 0.57 & 0.77 & 1.05 & 0.76 & -0.15 \\ -0.95 & 1.12 & -0.28 & 0.85 & 0.68 & 0.76 & 1.26 & 0.22 \\ 0.18 & 0.11 & 0.28 & 0.14 & -0.10 & -0.15 & 0.22 & 0.32 \end{array}\right) \end{array}.$$

Hence,

$$\mathcal{S}_{XX} = \begin{pmatrix} 1.41 & -1.11 \\ -1.11 & 1.19 \end{pmatrix}, \quad \mathcal{S}_{XY} = \begin{pmatrix} 0.78 & -0.71 & -0.90 & -1.04 & -0.95 & 0.18 \\ -0.42 & 0.82 & 0.77 & 0.90 & 1.12 & 0.11 \end{pmatrix},$$

$$\mathcal{S}_{YY} = \begin{pmatrix} 0.75 & -0.23 & -0.45 & -0.42 & -0.28 & 0.28 \\ -0.23 & 0.66 & 0.52 & 0.57 & 0.85 & 0.14 \\ -0.45 & 0.52 & 0.72 & 0.77 & 0.68 & -0.10 \\ -0.42 & 0.57 & 0.77 & 1.05 & 0.76 & -0.15 \\ -0.28 & 0.85 & 0.68 & 0.76 & 1.26 & 0.22 \\ 0.28 & 0.14 & -0.10 & -0.15 & 0.22 & 0.32 \end{pmatrix}.$$

It is interesting to see that value stability and price have a negative covariance. This makes sense since highly priced vehicles tend to loose their market value at a faster pace than medium priced vehicles.

Now we estimate $\mathcal{K} = \Sigma_{XX}^{-1/2}\, \Sigma_{XY}\, \Sigma_{YY}^{-1/2}$ by

$$\hat{\mathcal{K}} = \mathcal{S}_{XX}^{-1/2}\, \mathcal{S}_{XY}\, \mathcal{S}_{YY}^{-1/2}$$

and perform a singular value decomposition of $\hat{\mathcal{K}}$:

$$\hat{\mathcal{K}} = \mathcal{G}\mathcal{L}\mathcal{D}^{\top} = (g_1, g_2)\ \mathrm{diag}(\ell_1^{1/2}, \ell_2^{1/2})\ (d_1, d_2)^{\top}$$

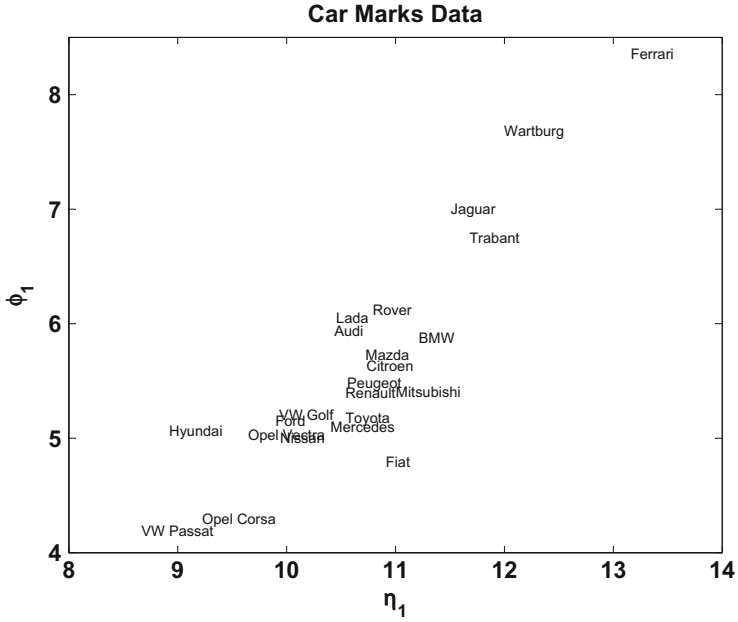

Fig. 16.1 The second canonical variables for the car marks data texttttMVAcancarm

where the ℓ_i's are the eigenvalues of $\hat{\mathcal{K}}\hat{\mathcal{K}}^\top$ and $\hat{\mathcal{K}}^\top\hat{\mathcal{K}}$ with $\text{rank}(\hat{\mathcal{K}}) = 2$, and g_i and d_i are the eigenvectors of $\hat{\mathcal{K}}\hat{\mathcal{K}}^\top$ and $\hat{\mathcal{K}}^\top\hat{\mathcal{K}}$, respectively. The canonical correlation coefficients are

$$r_1 = \ell_1^{1/2} = 0.98, \quad r_2 = \ell_2^{1/2} = 0.89.$$

The high correlation of the second two canonical variables can be seen in Fig. 16.1. The second canonical variables are

$$\hat{\eta}_1 = \hat{a}_1^\top x = 1.602\, x_1 + 1.686\, x_2$$

$$\hat{\varphi}_1 = \hat{b}_1^\top y = 0.568\, y_1 + 0.544\, y_2 - 0.012\, y_3 - 0.096\, y_4 - 0.014\, y_5 + 0.915\, y_6.$$

Note that the variables y_1 (economy), y_2 (service) and y_6 (easy handling) have positive coefficients on $\hat{\varphi}_1$. The variables y_3 (design), y_4 (sporty car) and y_5 (safety) have a negative influence on $\hat{\varphi}_1$.

The canonical variable η_1 may be interpreted as a price and value index. The canonical variable φ_1 is mainly formed from the qualitative variables economy, service and handling with negative weights on design, safety and sportiness. These variables may therefore be interpreted as an appreciation of the value of the car. The sportiness has a negative effect on the price and value index, as do the design and the safety features.

Testing the Canonical Correlation Coefficients

The hypothesis that the two sets of variables $\mathcal{X}$ and $\mathcal{Y}$ are uncorrelated may be tested (under normality assumptions) with Wilks likelihood ratio statistic (Gibbins, 1985):

$$T^{2/n} = \left|\mathcal{I} - S_{YY}^{-1} S_{YX} S_{XX}^{-1} S_{XY}\right| = \prod_{i=1}^{k} (1 - \ell_i).$$

This statistic unfortunately has a rather complicated distribution. Bartlett (1939) provides an approximation for large n:

$$-\{n - (p + q + 3)/2\} \log \prod_{i=1}^{k} (1 - \ell_i) \sim \chi^2_{pq}. \tag{16.14}$$

A test of the hypothesis that only s of the canonical correlation coefficients are nonzero may be based (asymptotically) on the statistic

$$-\{n - (p + q + 3)/2\} \log \prod_{i=s+1}^{k} (1 - \ell_i) \sim \chi^2_{(p-s)(q-s)}. \tag{16.15}$$

Example 16.2 Consider Example 16.1 again. There are $n = 40$ persons that have rated the cars according to different categories with $p = 2$ and $q = 6$. The canonical correlation coefficients were found to be $r_1 = 0.98$ and $r_2 = 0.89$. Bartlett's statistic (16.14) is therefore

$$-\{40 - (2 + 6 + 3)/2\} \log\{(1 - 0.98^2)(1 - 0.89^2)\} = 165.59 \sim \chi^2_{12}$$

which is highly significant (the 99 % quantile of the χ^2_{12} is 26.23). The hypothesis of no correlation between the variables $\mathcal{X}$ and $\mathcal{Y}$ is therefore rejected.

Let us now test whether the second canonical correlation coefficient is different from zero. We use Bartlett's statistic (16.15) with $s = 1$ and obtain

$$-\{40 - (2 + 6 + 3)/2\} \log\{(1 - 0.89^2)\} = 54.19 \sim \chi^2_5$$

which is again highly significant with the χ^2_5 distribution.

CCA with Qualitative Data

The canonical correlation technique may also be applied to qualitative data. Consider for example the contingency table $\mathcal{N}$ of the French baccalauréat data. The dataset is given in Table 22.8 in Chap. 22. The CCA cannot be applied directly to

this contingency table since the table does not correspond to the usual data matrix structure. We may wish, however, to explain the relationship between the row r and column c categories. It is possible to represent the data in a $(n \times (r+c))$ data matrix $\mathcal{Z} = (\mathcal{X}, \mathcal{Y})$ where n is the total number of frequencies in the contingency table $\mathcal{N}$ and $\mathcal{X}$ and $\mathcal{Y}$ are matrices of zero-one dummy variables. More precisely, let

$$x_{ki} = \begin{cases} 1 & \text{if the } k\text{th individual belongs to the } i\text{th row category} \\ 0 & \text{otherwise} \end{cases}$$

and

$$y_{kj} = \begin{cases} 1 & \text{if the } k\text{th individual belongs to the } j\text{th column category} \\ 0 & \text{otherwise} \end{cases}$$

where the indices range from $k = 1, \ldots, n$, $i = 1, \ldots, r$ and $j = 1, \ldots, c$. Denote the cell frequencies by n_{ij} so that $\mathcal{N} = (n_{ij})$ and note that

$$x_{(i)}^{\top} y_{(j)} = n_{ij},$$

where $x_{(i)}$ ($y_{(j)}$) denotes the ith (jth) column of $\mathcal{X}$ ($\mathcal{Y}$).

$$\mathcal{X} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \mathcal{Y} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \mathcal{Z} = (\mathcal{X}, \mathcal{Y}) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

Example 16.3 Consider the following example where

$$\mathcal{N} = \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}.$$

The matrices $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are therefore

The element n_{12} of $\mathcal{N}$ may be obtained by multiplying the first column of $\mathcal{X}$ with the second column of $\mathcal{Y}$ to yield

$$x_{(1)}^{\top} y_{(2)} = 2.$$

The purpose is to find the canonical variables $\eta = a^\top x$ and $\varphi = b^\top y$ that are maximally correlated. Note, however, that x has only one nonzero component and therefore an "individual" may be directly associated with its canonical variables or score (a_i, b_j). There will be n_{ij} points at each (a_i, b_j) and the correlation represented by these points may serve as a measure of dependence between the rows and columns of $\mathcal{N}$.

Let $\mathcal{Z} = (\mathcal{X}, \mathcal{Y})$ denote a data matrix constructed from a contingency table $\mathcal{N}$. Similar to Chap. 14 define

$$c = x_{i\bullet} = \sum_{j=1}^{c} n_{ij},$$

$$d = x_{\bullet j} = \sum_{i=1}^{r} n_{ij},$$

and define $\mathcal{C} = \operatorname{diag}(c)$ and $\mathcal{D} = \operatorname{diag}(d)$. Suppose that $x_{i\bullet} > 0$ and $x_{\bullet j} > 0$ for all i and j. It is not hard to see that

$$nS = \mathcal{Z}^\top \mathcal{H} \mathcal{Z} = \mathcal{Z}^\top \mathcal{Z} - n\bar{z}\bar{z}^\top = \begin{pmatrix} nS_{XX} & nS_{XY} \\ nS_{YX} & nS_{YY} \end{pmatrix}$$

$$= \left(\frac{n}{n-1}\right) \begin{pmatrix} \mathcal{C} - n^{-1} c c^\top & \mathcal{N} - \hat{\mathcal{N}} \\ \mathcal{N}^\top \hat{\mathcal{N}}^\top & \mathcal{D} - n^{-1} d d^\top \end{pmatrix}$$

where $\hat{\mathcal{N}} = cd^\top / n$ is the estimated value of $\mathcal{N}$ under the assumption of independence of the row and column categories.

Note that

$$(n-1) S_{XX} 1_r = \mathcal{C} 1_r - n^{-1} c c^\top 1_r = c - c(n^{-1} c^\top 1_r) = c - c(n^{-1} n) = 0$$

and therefore S_{XX}^{-1} does not exist. The same is true for S_{YY}^{-1}. One way out of this difficulty is to drop one column from both $\mathcal{X}$ and $\mathcal{Y}$, say the first column. Let $\bar{c}$ and $\bar{d}$ denote the vectors obtained by deleting the first component of c and d.

Define $\bar{\mathcal{C}}$, $\bar{\mathcal{D}}$ and $\bar{S}_{XX}$, $\bar{S}_{YY}$, $\bar{S}_{XY}$ accordingly and obtain

$$(n\bar{S}_{XX})^{-1} = \bar{\mathcal{C}}^{-1} + n_{i\bullet}^{-1} 1_r 1_r^\top$$

$$(n\bar{S}_{YY})^{-1} = \bar{\mathcal{D}}^{-1} + n_{\bullet j}^{-1} 1_c 1_c^\top$$

so that (16.3) exists. The score associated with an individual contained in the first row (column) category of $\mathcal{N}$ is 0.

The technique described here for purely qualitative data may also be used when the data is a mixture of qualitative and quantitative characteristics. One has to "blow up" the data matrix by dummy zero-one values for the qualitative data variables.

Summary

↪ In practice we estimate Σ_{XX}, Σ_{XY}, Σ_{YY} by the empirical covariances and use them to compute estimates ℓ_i, g_i, d_i for λ_i, γ_i, δ_i from the SVD of $\hat{\mathcal{K}} = \mathcal{S}_{XX}^{-1/2}\mathcal{S}_{XY}\mathcal{S}_{YY}^{-1/2}$.

↪ The signs of the coefficients of the canonical variables tell us the direction of the influence of these variables.

16.3 Exercises

Exercise 16.1 *Show that the eigenvalues of $\mathcal{K}\mathcal{K}^\top$ and $\mathcal{K}^\top\mathcal{K}$ are identical. (Hint: Use Theorem 2.6.)*

Exercise 16.2 *Perform the CCA for the following subsets of variables: $\mathcal{X}$ corresponding to {price} and $\mathcal{Y}$ corresponding to {economy, easy handling} from the car marks data (Table 22.7).*

Exercise 16.3 *Calculate the first canonical variables for Example 16.1. Interpret the coefficients.*

Exercise 16.4 *Use the SVD of matrix $\mathcal{K}$ to show that the canonical variables η_1 and η_2 are not correlated.*

Exercise 16.5 *Verify that the number of nonzero eigenvalues of matrix $\mathcal{K}$ is equal to* $\operatorname{rank}(\Sigma_{XY})$.

Exercise 16.6 *Express the singular value decomposition of matrices $\mathcal{K}$ and $\mathcal{K}^\top$ using eigenvalues and eigenvectors of matrices $\mathcal{K}^\top\mathcal{K}$ and $\mathcal{K}\mathcal{K}^\top$.*

Exercise 16.7 *What will be the result of CCA for $Y = X$?*

Exercise 16.8 *What will be the results of CCA for $Y = 2X$ and for $Y = -X$?*

Exercise 16.9 *What results do you expect if you perform CCA for X and Y such that $\Sigma_{XY} = 0$? What if $\Sigma_{XY} = \mathcal{I}_p$?*