Canonical Correlation Analysis

Consider a multivariate random vector of the form (x, y). Suppose we are given a sample of instances $S = ((x_1, y_1), ..., (x_n, y_n))$ of (x, y). Let S_x denote $(x_1, ..., x_n)$ and similarly S_y denote $(y_1, ..., y_n)$. We can consider defining a new coordinate for x by choosing a direction w_x and projecting x onto that direction,

$$x \to \langle w_x, x \rangle.$$

If we do the same for y by choosing a direction w_y, we obtain a sample of the new x coordinate. Let

$$S_{x,w_x} = (\langle w_x, x_1 \rangle, ..., \langle w_x, x_n \rangle),$$

with the corresponding values of the new y coordinate being

$$S_{y,w_y} = (\langle w_y, y_1 \rangle, ..., \langle w_y, y_n \rangle),$$

The first stage of canonical correlation is to choose w_x and w_y to maximize the correlation between the two vectors. In other words, the functions result to be maximized is

$$\begin{aligned}\rho &= \max_{w_x, w_y} corr(S_x w_x, S_y w_y) \\ &= \max_{w_x, w_y} \frac{\langle S_x w_x, S_y w_y \rangle}{||S_x w_x|| \, ||S_y w_y||}\end{aligned}$$

If we use $\hat{\mathbb{E}}[f(x, y)]$ to denote the empirical expectation of the function $f(x, y)$, where

$$\hat{\mathbb{E}}[f(x, y)] = \frac{1}{m} \sum_{i=1}^{m} f(x_i, y_i),$$

we can rewrite the correlation expression as

$$\begin{aligned}\rho &= \max_{w_x, w_y} \frac{\hat{\mathbb{E}}[\langle w_x, x \rangle \langle w_y, y \rangle]}{\sqrt{\hat{\mathbb{E}}[\langle w_x, x \rangle^2] \hat{\mathbb{E}}[\langle w_y, y \rangle^2]}} \\ &= \max_{w_x, w_y} \frac{\hat{\mathbb{E}}[w_x' x y' w_y]}{\sqrt{\hat{\mathbb{E}}[w_x' x x' w_x] \hat{\mathbb{E}}[w_y' y y' w_y]}}\end{aligned}$$

It follows that

$$\rho = \max_{w_x, w_y} \frac{w_x^{'} \hat{\mathbb{E}}[w_x^{'} x y^{'}] w_y}{\sqrt{w_x^{'} \hat{\mathbb{E}}[x x^{'}] w_x w_y^{'} \hat{\mathbb{E}}[y y^{'}] w_y}}$$

Now observe that the covariance matrix of (x, y) is

$$\mathrm{C}(x, y) = \hat{\mathbb{E}}\left[\begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}^{'}\right] = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} = C \tag{1}$$

The total covariance matrix C is a block matrix where the within-sets covariance matrices are C_{xx} and C_{yy} and the between-sets covariance matrices are $\mathrm{C}_{xy} = \mathrm{C}_{yx}^{'}$, although equation (1) is the covariance matrix only in the zero-mean case.
Hence, we can rewrite the function ρ as

$$\rho = \max_{w_x, w_y} \frac{w_x^{'} C_{xy} w_y}{\sqrt{w_x^{'} C_{xx} w_x w_y^{'} C_{yy} w_y}} \tag{2}$$

The maximum canonical correlation is the maximum of ρ with respect to w_x and w_y

.

Algorithm

In this section, we give an overview of the CCA algorithms where we formulate the optimization problem as a standard eigenproblem.

Observe that the solution of equation (2) is not affected by rescaling w_x or w_y either together or independently, so that, for example, replacing w_x by αw_x gives the quotient

$$\frac{\alpha w_x^{'} C_{xy} w_y}{\sqrt{\alpha^2 w_x^{'} C_{xx} w_x w_y^{'} C_{yy} w_y}} = \frac{w_x^{'} C_{xy} w_y}{\sqrt{w_x^{'} C_{xx} w_x w_y^{'} C_{yy} w_y}}$$

Since the choice of rescaling is therefore arbitrary, the CCA optimization problem formulated in equation (2) is equivalent to maximizing the numerator subject to

$$w_x^{'} C_{xx} w_x = 1$$
$$w_y^{'} C_{yy} w_y = 1$$

The corresponding Lagrangian is

$$L(\lambda, w_x, w_y) = w_x^{'} C_{xx} w_y - \frac{\lambda_x}{2}(w_x^{'} C_{xx} w_x - 1) - \frac{\lambda_y}{2}(w_y^{'} C_{yy} w_y - 1)$$

Taking derivatives in respect to w_x and w_y, we obtain

$$\frac{\partial L}{\partial w_x} = C_{xy} w_y - \lambda_x C_{xx} w_x = 0 \tag{3}$$

$$\frac{\partial L}{\partial w_y} = C_{yx} w_x - \lambda_y C_{yy} w_y = 0 \tag{4}$$

Subtracting $w_y^{'}$ times the second equation from $w_x^{'}$ times the first, we have

$$\begin{aligned} 0 &= w_x^{'} C_{xy} w_y - w_x^{'} \lambda_x C_{xx} w_x - w_y^{'} C_{yx} w_x + w_y^{'} \lambda_y C_{yy} w_y \\ &= \lambda_y w_y^{'} C_{yy} w_y - \lambda_x w_x^{'} C_{xx} w_x \end{aligned}$$

which together with the constraints implies that $\lambda_y \lambda_x = 0$, let$\lambda = \lambda_x = \lambda_y$. Assuming C_{yy} is invertible, we have

$$w_y = \frac{C_{yy}^{-1} C_{yx} w_x}{\lambda} \tag{5}$$

and so substituting in equation (3) gives

$$C_{xy} C_{yy}^{-1} C_{yx} w_x = \lambda^2 C_{xx} w_x \tag{6}$$

We are left with a generalized eigenproblem of the form $Ax = \lambda Bx$. We can therefore find the coordinate system that optimizes the correlation between corresponding coordinates by first solving for the generalized eigenvectors of equation (6) to obtain the sequence of w_xs and then using equation (5) to find the corresponding w_ys.

If C_{xx} is invertible, we are able to formulate equation (6) as a standard eigenproblem of the form $B^{-1}Ax = \lambda x$, although to ensure a symmetric standard eigenproblem, we do the following. As the covariance matrices C_{xx} and C_{yy} are symmetric positive definite, we are able to decompose them using a complete Cholesky decomposition,

$$C_{xx} = R_{xx} . R_{xx}^{'}$$

where R_{xx} is a lower triangular matrix. If we let $u_x = R_{xx}^{'} . w_x$, we are able to rewrite equation (6) as follows:

$$\begin{aligned} C_{xy} C_{yy}^{-1} C_{yx} R_{xx}^{-1'} u_x &= \lambda^2 R_{xx} u_x \\ R_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} R_{xx}^{-1'} u_x &= \lambda^2 u_x \end{aligned}$$

We are therefore left with a symmetric standard eigenproblem of the form $Ax = \lambda x$.