Canonical Correlation Analysis with Kernels

Florian Markowetz
Max-Planck-Institute for Molecular Genetics
– Computational Molecular Biology –
Berlin

Computational Diagnostics Group Seminar
2003 Mar 10

Overview

"Applied Kernel Generalized Canonical Correlation Analysis Explained"

1. **CCA**: Canonical Correlation Analysis;

2. **GCCA**: Generalized Canonical Correlation Analysis;

3. **KGCCA**: Kernel Generalized Canoncial Correlation Analysis;

Canonical Correlation Analysis - in words

CCA seeks to identify and quantify the associations between two sets of variables. It searches for linear combinations of the original variables having maximal correlation.

Further pairs of maximally correlated linear combinations are chosen such, that they are othogonal to those already identified.

The pairs of linear combinations are called canonical variables and their correlations the canonical correlations. The canonical correlations measure the strength of association between the two sets of variables.

CCA is closely related to other linear subspace methods like Principal Component Analysis, Partial Least Squares and Multivariate Linear Regression.

Canonical Correlation Analysis - in formulas

Data: m microarrays measuring N genes, organized in $N \times m$ matrix Z.

$$Z = [\ X \mid Y\ ] \quad \begin{cases} X: \ N \times p \ \text{ matrix} \\ Y: \ N \times q \ \text{ matrix} \end{cases}$$

Linear combinations of the variables X_i and Y_i:

$$U_a = a^\top X = \sum_1^p a_i X_i \qquad\qquad V_b = b^\top Y = \sum_1^q b_i Y_i$$

Correlation is defined as:

$$corr(U,V) = \frac{cov(U,V)}{\sqrt{var(U)\ var(V)}}$$

Stating the CCA problem

CCA is the solution of the optimization problem:

$$\begin{aligned} \underset{a,b}{\text{maximize}} \quad & corr(U_a, V_b) \\ \text{subject to} \quad & var(U_a) = var(V_b) = 1 \end{aligned}$$

The maximal value ρ is the first canonical correlation.

The canonical variates U_α, V_β are defined by:

$$(\alpha, \beta) = \underset{a,b}{\operatorname{argmax}} \mid corr(a^\top X, b^\top Y) \mid$$

Solving the optimization problem

First we decompose $cov(Z)$:

$$cov(Z) \;=\; Z^\top Z \;\equiv\; R_Z \;=\; \begin{pmatrix} R_{XX} & R_{XY} \\ R_{YX} & R_{YY} \end{pmatrix}$$

Solving the optimization problem by Lagrange method leads to an Eigenvalue equation:

$$B^{-1}A\,w \;=\; \rho w$$

$$A = \begin{pmatrix} 0 & R_{XY} \\ R_{YX} & 0 \end{pmatrix} \qquad B = \begin{pmatrix} R_{XX} & 0 \\ 0 & R_{YY} \end{pmatrix} \qquad w = \begin{pmatrix} a' \\ b' \end{pmatrix}$$

Relation to other linear subspace methods

	A	B
PCA	R_{XX}	$\mathbb{I}$
CCA	$\begin{pmatrix} 0 & R_{XY} \\ R_{YX} & 0 \end{pmatrix}$	$\begin{pmatrix} R_{XX} & 0 \\ 0 & R_{YY} \end{pmatrix}$
PLS	$\begin{pmatrix} 0 & R_{XY} \\ R_{YX} & 0 \end{pmatrix}$	$\begin{pmatrix} \mathbb{I} & 0 \\ 0 & \mathbb{I} \end{pmatrix}$
MLR	$\begin{pmatrix} 0 & R_{XY} \\ R_{YX} & 0 \end{pmatrix}$	$\begin{pmatrix} R_{XX} & 0 \\ 0 & \mathbb{I} \end{pmatrix}$

Generalized Canonical Correlation Analysis

How to deal with more than two sets?

Straightforward: maximize the sum of all pairwise correlations.

Using kernels: Combining several datasets by summing up kernel matrices.

Kernel Canonical Correlation Analysis

What is a kernel function?

A similarity measure like the inner product $\langle x, y \rangle = \sum x_i y_i$.

$$k : S \times S \to \mathbb{R}$$

Given a (nonlinear) map Φ into a (highdimensional) Feature space, we can define a kernel function by:

$$k(x_i, x_j) \;:=\; \langle \Phi(x_i), \Phi(x_j) \rangle$$

What is a kernel matrix?

A positive definite matrix which summarizes the similarities of all members of a set.

$$K_{ij} \;=\; k(x_i, x_j)$$

The Kernel Trick

Given an algorithm which is formulated in terms of an inner product, one can construct an alternative algorithm by replacing the inner product by an kernel function k.

This is used in SVMs and can also be applied to CCA. Useful, because it makes linear algorithms nonlinear and heterogeneous datasets can be combined.

Examples of kernels

Radial basis function kernels - a kernel for vectorial data.

$$k(x_i, x_j) \; = \; exp(\|x_i - x_j\|/\sigma)$$

The diffusion kernel - a kernel for graphs.
Given an undirected graph $\Gamma = (V, E)$ with adjacency matrix A. Let A_{i+} be the sum over the i-th row of A. We define the matrix H by

$$H \; = \; A - diag(A_{1+}, \ldots, A_{n+})$$

The diffusion kernel is defined by the positive definite matrix K_{diff}:

$$K_{diff} \; = \; \exp(cH) \; = \; \sum \frac{c^k}{k!} H^k$$

Kernel Canonical Correlation Analysis

Both ordinary and kernel CCA can be written as the solution of an Eigenvalue equation of the form

$$B^{-1}A\, w \;=\; \rho w\ .$$

Ordinary CCA

$$A = \begin{pmatrix} 0 & R_{XY} \\ R_{YX} & 0 \end{pmatrix} \qquad B = \begin{pmatrix} R_{XX} & 0 \\ 0 & R_{YY} \end{pmatrix} \qquad w = \begin{pmatrix} a' \\ b' \end{pmatrix}$$

Kernel CCA

$$A = \begin{pmatrix} 0 & K_X K_Y \\ K_Y K_X & 0 \end{pmatrix} \qquad B = \begin{pmatrix} K_X K_X & 0 \\ 0 & K_Y K_Y \end{pmatrix} \qquad w = \begin{pmatrix} a' \\ b' \end{pmatrix}$$