Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 2

Recall: For a sample of d_1- and d_2-dimensional data of size N, given as two data matrices $X \in \mathbb{R}^{d_1 \times N}$, $Y \in \mathbb{R}^{d_2 \times N}$ (assumed to be centered), canonical correlation analysis (CCA) finds a one-dimensional projection maximizing the cross-correlation for constant auto-correlation. The primal optimization problem is:

$$\begin{aligned} \text{Find } w_x \in \mathbb{R}^{d_1}, w_y \in \mathbb{R}^{d_2} \text{ maximizing } \quad & w_x^\top C_{xy} w_y \\ \text{subject to } \quad & w_x^\top C_{xx} w_x = 1 \\ & w_y^\top C_{yy} w_y = 1, \end{aligned} \tag{1}$$

where $C_{xx} = \frac{1}{N} X X^\top \in \mathbb{R}^{d_1 \times d_1}$ and $C_{yy} = \frac{1}{N} Y Y^\top \in \mathbb{R}^{d_2 \times d_2}$ are the auto-covariance matrices of X resp. Y, and $C_{xy} = \frac{1}{N} X Y^\top \in \mathbb{R}^{d_1 \times d_2}$ is the cross-covariance matrix of X and Y.

Exercise 1: Primal CCA $(10 + 5$ P)

We have seen in the lecture that a solution of the canonical correlation analysis can be found in some eigenvector of the generalized eigenvalue problem:

$$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \lambda \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix}$$

(a) *Show* that among all eigenvectors (w_x, w_y) the solution is the one associated to the highest eigenvalue.

(b) *Show* that if (w_x, w_y) is a solution, then $(-w_x, -w_y)$ is also a solution of the CCA problem.

Exercise 2: Dual CCA $(10 + 15 + 5 + 5$ P)

In this exercise, we would like to derive the dual optimization problem.

(a) *Show*, that it is always possible to find an optimal solution in the span of the data, that is,

$$w_x = X\alpha_x \ , \quad w_y = Y\alpha_y$$

with some coefficient vectors $\alpha_x \in \mathbb{R}^N$ and $\alpha_y \in \mathbb{R}^N$.

(b) *Show* that the solution of the dual optimization problem is found in an eigenvector of the generalized eigenvalue problem

$$\begin{bmatrix} 0 & A \cdot B \\ B \cdot A & 0 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \rho \cdot \begin{bmatrix} A^2 & 0 \\ 0 & B^2 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$$

where $A = X^\top X$ and $B = Y^\top Y$.

(c) *Show* that the solution of the dual is given by the eigenvector associated to the highest eigenvalue.

(d) *Show* how a solution to the original problem can be obtained from the solution of the generalized eigenvalue problem of the dual.

Exercise 3: CCA and Least Square Regression (20 P)

Consider some supervised dataset with the inputs stored in a matrix $X \in \mathbb{R}^{D \times N}$ and the targets stored in a vector $y \in \mathbb{R}^N$. We assume that both our inputs and targets are centered. The least squares regression optimization problem is:

$$\min_{v \in \mathbb{R}^D} \quad \|X^\top v - y\|^2$$

We would like to relate least square regression and CCA, specifically, their respective solutions v and (w_x, w_y).

(a) *Show* that if X and y are the two modalities of CCA (i.e. $X \in \mathbb{R}^{D \times N}$ and $y \in \mathbb{R}^{1 \times N}$), the first part of the solution of CCA (i.e. the vector w_x) is equivalent to the solution v of least square regression up to a scaling factor.

Exercise 4: Programming (30 P)

Download the programming files on ISIS and follow the instructions.

Exercise Sheet 2

1 Primal CCA

a) $$\begin{pmatrix} w_x^T & w_y^T \end{pmatrix} \begin{pmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix} = \lambda \begin{pmatrix} w_x & w_y \end{pmatrix} \begin{pmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} w_y^T C_{yx} & w_x^T C_{xy} \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix} = \lambda \begin{pmatrix} w_x^T C_{xx} & w_y C_{yy} \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix}$$

$$\Leftrightarrow w_y^T C_{yx} w_x + w_x^T C_{xy} w_y = \lambda (w_x^T C_{xx} w_x + w_y^T C_{yy} w_y)$$

$$\Leftrightarrow w_y^T C_{yx} w_x + w_x^T C_{xy} w_y = 2\lambda$$

$$\Leftrightarrow (C_{yx} w_x)^T w_y + w_x^T C_{xy} w_y = 2\lambda$$

$$\Leftrightarrow w_x^T C_{yx}^T w_y + w_x C_{xy} w_y = 2\lambda$$

$$\Leftrightarrow 2 w_x^T C_{xy} w_y = 2\lambda$$

$$\Leftrightarrow w_x^T C_{xy} w_y = \lambda \quad \rightarrow \quad \max w_x^T C_{xy} w_y = \max \lambda$$

b) maximize $w_x^T C_{xy} w_y$ with $w_x^T C_{xx} w_x = 1 = w_y^T C_{yy} w_y$

$w_x \rightarrow -w_x, \quad w_y \Rightarrow -w_y$:

$$(-w_x)^T C_{xy} (-w_y) \rightarrow w_x^T C_{xy} w_y$$
$$(-w_x)^T C_{xx} (-w_x) = 1 = w_x^T C_{xx} w_x$$
$$(-w_y)^T C_{yy} (-w_y) = 1 = w_y^T C_{yy} w_y$$

The maximization problem and the constraints stay the same

$\rightarrow$ w_x, w_y yields the same solution as $-w_x, -w_y$

2 Dual CCA

a) $w_x = S_x + n_x$, $w_y = S_y + n_y$

$\max (S_x + n_x)^T C_{xy} (S_y + n_y)$ st $w_x^T C_{xy} w_y = (S_x + n_x)^T X Y^T (S_y + n_y)$

$n_x + n_y$ have no effect. We can write S_x as $X\alpha_x$ and S_y as $Y\alpha_y$ → weighted combination of our data will give us dual information

b) $\begin{pmatrix} 0 & XY^T \\ YX^T & 0 \end{pmatrix} \begin{pmatrix} X\alpha_x \\ Y\alpha_y \end{pmatrix} = \lambda \begin{pmatrix} XX^T & 0 \\ 0 & YY^T \end{pmatrix} \begin{pmatrix} X\alpha_x \\ Y\alpha_y \end{pmatrix}$

$$= \begin{pmatrix} X^T & 0 \\ 0 & Y^T \end{pmatrix} \begin{pmatrix} 0 & XY^T \\ YX^T & 0 \end{pmatrix} \begin{pmatrix} X\alpha_x \\ Y\alpha_y \end{pmatrix} = \lambda \begin{pmatrix} X^T & 0 \\ 0 & Y^T \end{pmatrix} \begin{pmatrix} XX^T & 0 \\ 0 & YY^T \end{pmatrix} \begin{pmatrix} X\alpha_x \\ Y\alpha_y \end{pmatrix}$$

$$= \begin{pmatrix} 0 & X^TXY^TY \\ Y^TYX^TX & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \lambda \begin{pmatrix} X^TXX^TX & 0 \\ 0 & Y^TYY^TY \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix}$$

$$= \begin{pmatrix} 0 & AB \\ BA & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \lambda \begin{pmatrix} AA & 0 \\ 0 & BB \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix}$$

~~Be~~ Better:

$$\begin{pmatrix} 0 & C_{xy}Y \\ C_{yx}X & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \lambda \begin{pmatrix} C_{xx}X & 0 \\ 0 & C_{yy}Y \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} 0 & XY^TY \\ YX^TX & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \lambda \begin{pmatrix} XX^TX & 0 \\ 0 & YY^TY \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} X^T & 0 \\ 0 & Y^T \end{pmatrix} \begin{pmatrix} 0 & XY^TY \\ YX^TX & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \lambda \begin{pmatrix} X^T & 0 \\ 0 & Y^T \end{pmatrix} \begin{pmatrix} XX^TX & 0 \\ 0 & YY^TY \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} 0 & X^TXY^TY \\ Y^TYX^TX & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \lambda \begin{pmatrix} X^TXX^TX & 0 \\ 0 & Y^TYY^TY \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} 0 & AB \\ BA & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \lambda \begin{pmatrix} A^2 & 0 \\ 0 & B^2 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix}$$

c) $\begin{pmatrix} 0 & AB \\ BA & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \rho \begin{pmatrix} A^2 & 0 \\ 0 & B^2 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix}$

$\Leftrightarrow \begin{pmatrix} 0 & X^TXY^TY \\ Y^TYX^T\cancel{X} & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \rho \begin{pmatrix} X^TXX^TX & 0 \\ 0 & Y^TYY^TY \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix}$

$\Leftrightarrow \begin{pmatrix} X^TXY^TY\alpha_y \\ Y^TYX^TX\alpha_x \end{pmatrix} = \rho \begin{pmatrix} X^TXX^TX\alpha_x \\ Y^TYY^TY\alpha_y \end{pmatrix}$

$\Leftrightarrow \begin{pmatrix} X^TXY^Tw_y \\ Y^TYX^Tw_x \end{pmatrix} = \rho \begin{pmatrix} X^TXX^Tw_y \\ Y^TYY^Tw_y \end{pmatrix}$

$\Leftrightarrow (\alpha_x^T \quad \alpha_y^T) \begin{pmatrix} X^TC_{xy}w_y \\ Y^TC_{yx}w_x \end{pmatrix} = \rho (\alpha_x^T \quad \alpha_y^T) \begin{pmatrix} X^TC_{xx}w_x \\ Y^TC_{yy}w_y \end{pmatrix}$

$\Leftrightarrow w_x^TC_{xy}w_y + w_y^TC_{yx}w_x = \rho(w_x^TC_{xx}w_x + w_y^TC_{yy}w_y)$

$\Leftrightarrow 2w_x^TC_{xy}w_y = 2\rho \Leftrightarrow w_x^TC_{xy}w_y = \rho$ → pick EV with highest EV.

d) $\begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} \rightarrow \begin{pmatrix} w_x \\ w_y \end{pmatrix} = \begin{pmatrix} X\alpha_x \\ Y\alpha_y \end{pmatrix}$

1) $\begin{pmatrix} 0 & AB \\ BA & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \lambda \begin{pmatrix} A^2 & 0 \\ 0 & B^2 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix}$

2) $\begin{pmatrix} 0 & AB \\ BA & 0 \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix} = \lambda \begin{pmatrix} A^2 & 0 \\ 0 & B^2 \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix}$

$\begin{pmatrix} 0 & AB \\ BA & 0 \end{pmatrix} \begin{pmatrix} X\alpha_x \\ Y\alpha_y \end{pmatrix} = \lambda \begin{pmatrix} A^2 & 0 \\ 0 & B^2 \end{pmatrix} \begin{pmatrix} X\alpha_x \\ Y\alpha_y \end{pmatrix}$

3 CCA and Least Square Regression

a) $\min_v \|X^T v - y\|^2 = \min_v (X^T v - y)^T (X^T v - y)$

$= \min_v (v^T X - y)(X^T v - y) = \min_v v^T X X^T v - v^T X y - y X^T v + y^2$

$= \min_v v^T X X^T v - v^T X y - (Xy)^T V = \min v^T X X^T v - 2v^T X y + y^2$

Find minimum: $\frac{\partial}{\partial v} v^T X X^T v - 2v^T X y = 0$

$\Leftrightarrow \frac{\partial}{\partial b} a^T X X^T b + \frac{\partial}{\partial a} a^T X X^T b + \frac{\partial}{\partial v} - 2v^T X y = 0$

$\Leftrightarrow v^T X X^T + v^T X X^T - 2y^T X^T = 0$

$\Leftrightarrow 2(XX^T)^T v - 2Xy = 0 \Leftrightarrow 2XX^T v - 2Xy = 0$

$\Leftrightarrow XX^T v = Xy \Leftrightarrow v = (XX^T)^{-1} X y$

From 1), $C_{xy} w_y = \lambda C_{xx} w_x \Leftrightarrow X_y w_y = \lambda X X^T w_x$

$\Leftrightarrow w_x = \frac{w_y}{\lambda} (XX^T)^{-1} X_y \Leftrightarrow w_x = \frac{1}{\lambda} w_y V$