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:

Find
$$w_x \in \mathbb{R}^{d_1}, w_y \in \mathbb{R}^{d_2}$$
 maximizing $w_x^\top C_{xy} w_y$
subject to $w_x^\top C_{xx} w_x = 1$
 $w_y^\top C_{yy} w_y = 1,$ (1)

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

Exercise 1: Primal CCA (15 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 (35 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. Least squares regression optimization problem is:

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

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.