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

Exercise 1: Dual CCA (40 P)

In this exercise, we would like to derive the dual optimization problem, which is more efficient to solve if $N \ll d_i$.

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

$$w_x = X\alpha_x, w_y = Y\alpha_y. \tag{2}$$

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

(b) *Show* that the dual optimization problem is equivalent to finding the solution of the generalized eigenvalue problem

$$\begin{bmatrix} 0 & A \cdot B \\ B \cdot A & 0 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \rho \begin{bmatrix} A^2 & 0 \\ 0 & B^2 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}, \tag{3}$$

in α_x, α_y, where $A = X^\top X$ and $B = Y^\top Y$.

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

Exercise 2: Kernelized CCA (30 P)

We consider a kernel $k_x(x, x')$ and $k_y(y, y')$ for each modality of the data. We denote by $K_X \in \mathbb{R}^{N \times N}$ and $K_Y \in \mathbb{R}^{N \times N}$ the Gram matrices of the respective kernels.

(a) *Describe* how the CCA problem and its generalized eigenvalue formulation can be kernelized.

(b) *Explain* how the solution of the resulting kernelized CCA are to be interpreted, and under which condition the solution can/cannot be expressed as directions in the input spaces $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$.

Exercise 3: Deep CCA (30 P)

We would like to perform CCA on the top layer representation of two neural networks. Abstracting each network as a feature map $\phi_x(x; \theta_x) \in \mathbb{R}^{h_1}$ and $\phi_y(y; \theta_y) \in \mathbb{R}^{h_2}$, with respective parameter vectors θ_x, θ_y, and where x, y are the two input modalities, we consider a relaxed and unconstrained form of CCA given by:

$$\max_{\theta_x, \theta_y, w_x, w_y} w_x^\top C_{xy} w_y + \alpha \cdot \left[\min\left(0\,,\,1 - w_x^\top C_{xx} w_x\right) + \min\left(0\,,\,1 - w_y^\top C_{yy} w_y\right)\right],$$

for $w_x \in \mathbb{R}^{h_1}$ and $w_y \in \mathbb{R}^{h_2}$, where the covariance matrices are defined as

$$C_{xy} = \mathbb{E}[\phi_x \phi_y^\top], \quad C_{xx} = \mathbb{E}[\phi_x \phi_x^\top], \quad C_{yy} = \mathbb{E}[\phi_y \phi_y^\top],$$

where the parameter $\alpha > 0$ is chosen at hand, and $\mathbb{E}[\cdot]$ is the expectation with respect to the input distribution.

(a) *Explain* how the unconstrained objective above relates to the original CCA objective.

(b) *Express* the gradient of the new objective with respect to θ_x as a function of the Jacobian matrix $\frac{\partial \phi_x}{\partial \theta_x}$.