# Machine Learning 2 – Group ESHG
# Assignment 02

Willi Gierke, Arik Elimelech, Mehmed Halilovic, Leon Sixt

May 3, 2017

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 $\alpha_x, \alpha_y$, 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 $\theta_x, \theta_y$, 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

# Task 1

## a

We start from the generalised eigenvalue equation as in the lecture:

$$
\begin{aligned}
\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} &= \alpha \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} \\
\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix}^{-1} \begin{bmatrix} w_x \\ w_y \end{bmatrix} &= \alpha \begin{bmatrix} w_x \\ w_y \end{bmatrix} \\
\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} C_{xx}^{-1} & 0 \\ 0 & C_{yy}^{-1} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} &= \alpha \begin{bmatrix} w_x \\ w_y \end{bmatrix} \\
\begin{bmatrix} 0 & C_{xy} C_{yy}^{-1} \\ C_{yx} C_{xx}^{-y} & 0 \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} &= \alpha \begin{bmatrix} w_x \\ w_y \end{bmatrix} \\
\begin{bmatrix} 0 & \frac{1}{N} X Y^T (\frac{1}{N} Y Y^T)^{-1} \\ \frac{1}{N} Y X^T (\frac{1}{N} X X^T)^{-1} & 0 \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} &= \alpha \begin{bmatrix} w_x \\ w_y \end{bmatrix}
\end{aligned}
$$

## b

The solution is given as general eigenvalue problem (from the slides):

$$
\begin{pmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix} = \alpha \begin{pmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix} \tag{1}
$$

Using data matricies and renaming $\alpha$ to $\rho$

$$
\begin{pmatrix} 0 & XY^T \\ YX^T & 0 \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix} = \alpha \begin{pmatrix} XX^T & 0 \\ 0 & YY^T \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix} \tag{2}
$$

Substituting $w_x = X\alpha_x$ and $w_y = Y\alpha_y$:

$$
\begin{pmatrix} 0 & XY^T \\ YX^T & 0 \end{pmatrix} \begin{pmatrix} X\alpha_x \\ Y\alpha_y \end{pmatrix} = \alpha \begin{pmatrix} XX^T & 0 \\ 0 & YY^T \end{pmatrix} \begin{pmatrix} X\alpha_x \\ Y\alpha_y \end{pmatrix} \tag{3}
$$

Now, we can move the data matricies into the block matricies on both sides:

$$
\begin{pmatrix} 0 & XY^TY \\ YX^TX & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \alpha \begin{pmatrix} XX^TX & 0 \\ 0 & YY^TY \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} \tag{4}
$$

Multipling both sides with $\left(X^T Y^T\right)$ yields:

$$
\begin{pmatrix} 0 & X^TXY^TY \\ Y^TYX^TX & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \alpha \begin{pmatrix} X^TXX^TX & 0 \\ 0 & Y^TYY^TY \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} \tag{5}
$$

Using $A = X^TX$ and $B = Y^TY$ gives the required form:

$$
\begin{pmatrix} 0 & AB \\ BA & 0 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} = \alpha \begin{pmatrix} A^2 & 0 \\ 0 & B^2 \end{pmatrix} \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} \tag{6}
$$

### c

The original solution can be recovered by:

$$w_x = X\alpha_x \tag{7}$$
$$w_y = Y\alpha_y \tag{8}$$

## Task 2

### a

The original CCA problem is given by:

$$\max_{w_x, w_y} w_x^T C_{xy} w_y \tag{9}$$
$$\text{s.t. } w_x^T C_{xx} w_x \tag{10}$$
$$w_y^T C_{yy} w_y \tag{11}$$
$$\tag{12}$$

Substiting the covariance matricies with the corresponding data matricies.

$$\max_{w_x, w_y} w_x^T X Y^T w_y \tag{13}$$
$$\text{s.t. } w_x^T X X^T w_x \tag{14}$$
$$w_y^T Y Y^T w_y \tag{15}$$
$$\tag{16}$$

As argued in exercise 1, we can replace $w_x = X\alpha_x$ $w_y = Y\alpha_y$. This yields:

$$\max_{\alpha_x, \alpha_y} \alpha_x^T X^T X Y^T Y \alpha_y \tag{17}$$
$$\text{s.t. } \alpha_x^T X^T X X^T X \alpha_x \tag{18}$$
$$\alpha_y^T Y^T Y Y^T Y \alpha_y \tag{19}$$
$$\tag{20}$$

Substituting $X^T X$ with the Gram matrix $K_X$ and analogously for $K_Y$, yields the kernilized CCA problem.

$$\max_{\alpha_x, \alpha_y} \alpha_x^T K_X^T K_Y^T \alpha_y \tag{21}$$
$$\text{s.t. } \alpha_x^T K_X^T K_X^T \alpha_x \tag{22}$$
$$\alpha_y^T K_Y^T K_Y^T \alpha_y \tag{23}$$
$$\tag{24}$$

The generalized eigenvalue problem stays the same except that $A = K_X$ and $B = K_Y$.

### b

## Task 3

### i

The unconstrained form allows to learn more highly correlated representations, which aren't supposed to be linear. Thus, a kernel for KCCA is learned while the mapping function is not as restricted as in KCCA.

### ii

$$\frac{\partial}{\partial \theta_x} \max_{\theta_x, \theta_y, w_x, w_y} w_x^T C_{xy} w_y + \alpha[\min(0, 1 - w_x^T C_{xx} w_x) + \min(0, 1 - w_y^T C_{yy} w_y)]$$
$$=$$