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 $C_{xx} = \frac{1}{N}XX^\top \in \mathbb{R}^{d_1 \times d_1}$ 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: 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: $(d_x+d_y) \times (d_x+d_2)$

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

$Av = \lambda B v$

$\Leftrightarrow B^{-1}Av = \lambda v$

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

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

$$w_x^\top C_{xy} w_y + w_y^\top C_{yx} w_x = \lambda \left(w_x^\top C_{xx} w_x + w_y^\top C_{yy} w_y \right)$$

$$w_x^\top C_{xy} w_y = \lambda$$

(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: Kernel CCA (10 + 15 + 5 + 5 P)

In this exercise, we would like to kernelize CCA.

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

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

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

$w_x = s_x + n_x, \quad w_y = s_y + n_y$ $\qquad s_x^\top X \neq 0$, $\quad n_x^\top X = 0$

$$\begin{aligned} w_x^\top C_{xy} w_y &= (s_x + n_x)^\top XY^\top (s_y + n_y) \\ &= s_x^\top XY^\top s_y + \underbrace{s_x^\top XY^\top n_y + n_x^\top XY^\top s_y + n_x^\top XY^\top n_y}_{=0} \end{aligned}$$

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

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

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

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

$$(*) \Leftrightarrow \begin{bmatrix} X^\top & 0 \\ 0 & Y^\top \end{bmatrix}\begin{bmatrix} 0 & XY^\top \\ YX^\top & 0 \end{bmatrix}\begin{bmatrix} X\alpha_x \\ Y\alpha_y \end{bmatrix} = \rho \begin{bmatrix} X^\top & 0 \\ 0 & Y^\top \end{bmatrix}\begin{bmatrix} XX^\top & 0 \\ 0 & YY^\top \end{bmatrix}\begin{bmatrix} X\alpha_x \\ Y\alpha_y \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} 0 & X^\top XY^\top Y \\ Y^\top YX^\top X & 0 \end{bmatrix}\begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \rho \begin{bmatrix} X^\top XX^\top X & 0 \\ 0 & Y^\top YY^\top Y \end{bmatrix}\begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$$

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

$$[\alpha_x^\top\ \alpha_y^\top]\begin{bmatrix} X^\top & 0 \\ 0 & Y^\top \end{bmatrix}\begin{bmatrix} 0 & XY^\top \\ YX^\top & 0 \end{bmatrix}\begin{bmatrix} X\alpha_x \\ Y\alpha_y \end{bmatrix} = \rho [\alpha_x^\top\ \alpha_y^\top]\begin{bmatrix} X^\top & 0 \\ 0 & Y^\top \end{bmatrix}\begin{bmatrix} XX^\top & 0 \\ 0 & YY^\top \end{bmatrix}\begin{bmatrix} X\alpha_x \\ Y\alpha_y \end{bmatrix}$$

$$[w_x^\top\ w_y^\top]\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix}\begin{bmatrix} w_x \\ w_y \end{bmatrix} = \rho\, [w_x^\top\ w_y^\top]\begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix}\begin{bmatrix} w_x \\ w_y \end{bmatrix}$$

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

$w_x = X\alpha_x \qquad w_y = Y\alpha_y \qquad k(x, x') = \exp(-\gamma\|x - x'\|^2)$

$x \to \phi(x) \qquad x^\top w_x \to x^\top X\alpha_x \to k(x, X)\,\alpha_x$

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:

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

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.

$$\min_v\ (v^\top X - y^\top)(X^\top v - y) = \min_v\ v^\top XX^\top v - 2v^\top Xy = J$$

$$\frac{\partial J}{\partial v} = 2XX^\top v - 2Xy \overset{!}{=} 0$$

$$\Leftrightarrow v = (XX^\top)^{-1}Xy$$

$$C_{xy} w_y = \lambda C_{xx} w_x$$

$$\Leftrightarrow XY^\top w_y = \lambda XX^\top w_x$$

$$\Leftrightarrow w_x = (XX^\top)^{-1}XY^\top \cdot \frac{1}{\lambda} w_y \quad \text{scalar}$$