Exercises for the course
Machine Learning for Data Science
Winter Semester 2024/25

G. Montavon
Institute of Computer Science
Department of Mathematics and Computer Science
Freie Universität Berlin

Exercise Sheet 6 (theory part)

Exercise 1: Canonical Correlation Analysis (15 + 5 P)

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 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

$$\begin{aligned} C_{xx} &= \tfrac{1}{N} X X^\top \in \mathbb{R}^{d_1 \times d_1} \qquad \text{and} \\ C_{yy} &= \tfrac{1}{N} Y Y^\top \in \mathbb{R}^{d_2 \times d_2} \end{aligned}$$

are the auto-covariance matrices of X resp. Y, and

$$C_{xy} = \tfrac{1}{N} X Y^\top \in \mathbb{R}^{d_1 \times d_2}$$

is the cross-covariance matrix of X and Y. We also define $C_{yx} = \frac{1}{N} Y X^\top = C_{xy}^\top$.

(a) *Show* that a solution of the canonical correlation analysis can be found in some eigenvector of the generalized eigenvalue problem:

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

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

Exercise 2: CCA for High Dimensional Data (10 + 10 + 5 + 5 P)

Like for PCA the original problem formulation involving the eigendecomposition of $d \times d$ convariance matrices does not scale well for high-dimensional data. Here, we would like to derive another formulation of the CCA problem that involves instead the eigendecomposition a matrix whose size scales with the number of data points.

(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 resulting optimization problem is found in an eigenvector of the generalized eigenvalue problem

$$\begin{bmatrix} 0 & Q_{xy} \\ Q_{yx} & 0 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \lambda \begin{bmatrix} Q_{xx} & 0 \\ 0 & Q_{yy} \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$$

where $Q_{xy} = \frac{1}{N} X^\top X Y^\top Y$, $Q_{xx} = \frac{1}{N} X^\top X X^\top X$ and $Q_{yy} = \frac{1}{N} Y^\top Y Y^\top Y$.

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

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