Übung zur Vorlesung

Maschinelles Lernen 2

Sommersemester 2013

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 4

Due May 13, 2013, at 9:00 a.m. local time in electronic form via the ISIS website. See the corresponding submission page for formatting details (.zip archives, source code without fancy non-ASCII symbols etc.)

6. **Kernel Canonical Correlation Analysis (50 + 20 P)**
(a) Recall: For a sample of d_1- and d_2-dimensional data of size N, given as two data matrices $X \in \mathbb{R}^{d_1 \times N}$, $Y \in \mathbb{R}^{d_2 \times N}$, 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.

Derive the dual optimization problem, which is more efficient to solve if $N << d_i$. First, *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 $\alpha_y \in \mathbb{R}^N$. Then, *show* that the dual optimization problem is equivalent to finding the solution of the generalized eigenvalue problem

$$\begin{bmatrix} 0 & K_X K_Y \\ K_Y K_X & 0 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \rho \begin{bmatrix} K_X^2 & 0 \\ 0 & K_Y^2 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}, \tag{3}$$

in α_x, α_y, where $K_X = X^\top X$ and $K_Y = Y^\top Y$ are the *linear kernel matrices* of the data.

(b) *Describe* how the generalized eigenvalue problem from exercise (a) - and thus CCA - can be kernelized.

7. **tkCCA (30 P)**
On ISIS, you can find a MATLAB implementation of the tkCCA algorithm. Use this algorithm on the provided data set to find a temporal correlation. The data set consists of two time series $x \in \mathbb{R}^{20}$ and $y \in \mathbb{R}^{30}$. The data set contains a hidden one-dimensional signal which occurs in x without delay, and in y with delay τ. The script

```
>> tkcca_example
```

performs tkCCA between x and y, and then plots the results. Use the *canonical correlogram* to find the optimal τ, i.e., the time delay in y which maximizes the cross-correlation between x and y.

Please ask questions in the ISIS discussion forums for Machine Learning 2:
https://www.isis.tu-berlin.de/course/view.php?id=8005