Machine learning 2
Exercise sheet 4

FLEISCHMANN Kay, Matrnr: 352247
ROHRMANN Till, Matrnr: 343756

May 7, 2013

6 Kernel Canonical Correlation Analysis

(a) Given some trainingdata $X \in \mathbb{R}^{d1 \times N}$ and $Y \in \mathbb{R}^{d2 \times N}$. The idea behind CCA is to find two one dimensional projections w_x and w_y into space in which a latent variable exists with maximal correlation. Let $C_{xx} = XX^T$, $C_{yy} = YY^T$, $C_{xy} = XY^T$ and $C_{yx} = YX^T$.

Formally, Find $w_x \in \mathbb{R}^{d1}$, $w_y \in \mathbb{R}^{d2}$ which

$$maximaize \quad w_x^T C_{xy} w_y \tag{1}$$

with subject to

$$w_x^T C_{xx} w_x = 1 \tag{2}$$
$$w_y^T C_{yy} w_y = 1 \tag{3}$$

Show that it is always possible to find an optimal solution in the span of the data, that is $w_x = X\alpha_x$, $w_y = Y\alpha_y$:

$$TODO?? \tag{4}$$

Derive the dual optimization problem:

$$\mathcal{L}(\alpha, \beta) \quad = \quad w_x^T C_{xy} w_y - \frac{1}{2}\alpha(w_x^T C_{xx} w_x - 1)\frac{1}{2}\beta(w_y^T C_{yy} w_y - 1) \tag{5}$$

$$\frac{\partial L}{\partial w_x^T} = XY^T w_y - \alpha(XX^T w_x) = 0 \Leftrightarrow XY^T w_y = \alpha(XX^T w_x) \tag{6}$$
$$\frac{\partial L}{\partial w_y^T} = XY^T w_x - \beta(YY^T w_y) = 0 \Leftrightarrow XY^T w_x = \beta(YY^T w_y) \tag{7}$$

Multiplication w_x^T with Equation (6) and w_y^T with Equation (7) results to:

$$w_x^T XY^T w_y = \alpha(w_x^T XX^T w_x) \tag{8}$$
$$w_y^T XY^T w_x = \beta(w_y^T YY^T w_y) \tag{9}$$

Because of constraints (1) and (2)

$$\alpha(XX^T w_x) = \beta(w_y^T YY^T w_y) \Rightarrow \alpha = \beta \tag{10}$$

Next we combine equation (6)-(10)

$$C_{xy} w_y = \alpha C_{xx} w_x \tag{11}$$
$$C_{yx} w_x = \alpha C_{yy} w_y \tag{12}$$

and witen in a matrix we get finally an eigenvalue problem:

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

6.1 Kernel

With the assumtion of (4) and the equations (1)-(3) and $w_x = X\alpha_x$, $w_y = Y\alpha_y$:

$$maximize \quad \alpha_x^T X^T X Y^T Y \alpha_y = \alpha_x^T K_x K_y \alpha_y \tag{13}$$

with subject to

$$\alpha_x^T X^T X X^T X \alpha_x = \alpha_x^T K_x K_x \alpha_x = 1 \tag{14}$$
$$\alpha_y^T Y^T Y Y^T Y \alpha_y = \alpha_y^T K_y K_y \alpha_y = 1 \tag{15}$$
$$\tag{16}$$

Put this into the matrix formulation as followed in (12)
Kernel-equation:
$$\begin{bmatrix} 0 & K_x K_y \\ K_y K_x & 0 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \rho \begin{bmatrix} K_x^2 & 0 \\ 0 & K_y^2 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$$

(b) *Describe how the generalized eiganvalue problem from exercise (a) - and thun CCA - can be kernalized.* Computing the covariance of $X^T X$ or $Y^T Y$ for realy big matrices can be very expensive. The Kernel-Equation shows, that the solution of the generalized eigenvalue problem just depends on the scalar product between the datapoints. Therefore knowlege about the kernel (this means the scalarproduct between points $X^T X$ or $Y^T Y$) is enough to compute valid solutions.