

Blatt 4

Abgabe bis **Montag, 12. Mai 2010, 13 Uhr** bei Dr. Konrad Rieck per email
 (konrad.rieck@tu-berlin.de)

In diesem Aufgabenblatt werden Canonical Correlation Analysis (CCA) und die Erweiterungen kernel CCA (kCCA) und temporal kernel CCA (tkCCA) untersucht.

1. (**Kernel Canonical Correlation Analysis, 25 Punkte**) Gegeben seien zwei Datenmatrizen $X \in \mathbb{R}^{U \times N}$, $Y \in \mathbb{R}^{V \times N}$ (N sei die Anzahl der Datenpunkte, U , V die Dimensionalität der jeweiligen Variable). Die CCA Lösung maximiert die Kreuz-Korrelation und beschränkt die Auto-Korrelation in jeder Variable auf 1:

$$\begin{aligned} \operatorname{argmax}_{w_x, w_y} \quad & w_x^\top C_{xy} w_y \\ \text{s.t.} \quad & w_x^\top C_{xx} w_x = 1 \\ & w_y^\top C_{yy} w_y = 1. \end{aligned} \quad (1)$$

Hier bezeichnen $C_{xx} = XX^\top \in \mathbb{R}^{U \times U}$ und $C_{yy} = YY^\top \in \mathbb{R}^{V \times V}$ die Auto-Kovarianz Matrizen und $C_{xy} = XY^\top \in \mathbb{R}^{U \times V}$ die Kreuz-Kovarianz Matrix zwischen x und y . Bei *Kernel CCA* werden die Lösungen $w_x \in \mathbb{R}^U$, $w_y \in \mathbb{R}^V$ als Linearkombination von Datenpunkten dargestellt:

$$w_x = X\alpha, w_y = Y\beta. \quad (2)$$

Zeigen Sie, dass die Koeffizienten α , β gegeben sind durch die Lösung der generalisierten Eigenwertgleichung

$$\begin{bmatrix} 0 & K_X K_Y \\ K_Y K_X & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \rho \begin{bmatrix} K_X^2 & 0 \\ 0 & K_Y^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad (3)$$

wobei $K_X = X^\top X$ und $K_Y = Y^\top Y$ die *linearen Kerne* der Datenmatrizen bezeichnen.

2. (**Anwendung, 5 Punkte**) Auf der Website finden Sie einen Datensatz und eine Matlab Implementation des tkCCA Algorithmus. Im Datensatz finden Sie eine Variable $x \in \mathbb{R}^{20}$ und $y \in \mathbb{R}^{30}$. In einer Dimension von x, y findet sich ein verstecktes Signal, das in x ohne Zeitverzögerung und in y mit einer Zeitverzögerung τ gemessen wurde. Wenn Sie in der MATLAB Konsole

```
>> tkcca_example
```

aufrufen, wird tkCCA zwischen x und y berechnet und das Ergebnis geplottet. Bestimmen Sie anhand des *canonical correlograms* das optimale τ , d.h. die Anzahl der Samples um die x nach vorne (y nach hinten) verschoben werden muss, um die Korrelation zwischen x und y zu maximieren.