

Übungen zur Vorlesung  
**Wissensentdeckung in Datenbanken**  
Sommersemester 2010

Blatt 10

**Aufgabe 10.1 (4 Punkte)**

In dieser Aufgabe sollen Sie für  $\vec{x} = (x_1, \dots, x_p)$  zwei Funktionen  $f_1(\vec{x})$  und  $f_2(\vec{x})$  unter Nebenbedingungen  $h_i(\vec{x})$  optimieren. Stellen Sie dazu, wie auf den Folien, die entsprechende Lagrange-Funktion  $L(\vec{x}, \vec{\mu})$  auf. Die Lösung ergibt sich aus dem Gleichungssystem, das durch Nullsetzen der partiellen Ableitungen von  $L$  nach  $\vec{x}$  und  $\vec{\mu}$  entsteht.

- (a) Maximiere  $f_1(\vec{x}) = 1 - x_1^2 - x_2^2$  unter der Nebenbedingung  $h(\vec{x}) = x_1 + x_2 - 1 = 0$ .
- (b) Maximiere  $f_2(\vec{x}) = -x_1 - x_2 - x_3$  unter den Nebenbedingungen  $h_1(\vec{x}) = x_1^2 + x_2 - 3 = 0$  und  $h_2(\vec{x}) = x_1 + 3x_2 + 2x_3 - 7 = 0$ .

*Hinweis:* Beachten Sie bitte, dass für die bei Gleichheitsbedingungen  $h_i(\vec{x}) = 0$  eingeführten Multiplikatoren  $\mu_i$  nur  $\mu_i \neq 0$  (nicht  $\mu_i \geq 0$ !) gelten muss.

**Aufgabe 10.2 (2 Punkte)**

Sie haben die strukturelle Risikominimierung kennengelernt, die dazu dient, die Ausdrucksstärke von Modellklassen zu messen. Sie basiert auf der VC-Dimension.

- (a) Das empirische Risiko zweier jeweils bester Modelle aus unterschiedlichen Modellklassen sei identisch. Die erste Modellklasse besitze endliche, die zweite Modellklasse unendliche VC-Dimension. Welches Modell empfiehlt sich hinsichtlich der strukturellen Risikominimierung und warum?
- (b) Sei  $\mathcal{APR}_2$  die Klasse der achsenparallelen Rechtecke in der Ebene (also im  $\mathbb{R}^2$ ). Beweisen Sie (z. B. graphisch), dass diese Hypothesenklasse eine VC-Dimension von 4 besitzt. Dabei klassifiziert jedes Modell genau die Punkte innerhalb des Rechtecks als positiv.

**Aufgabe 10.3 (2 Punkte)**

In der Vorlesung wurde zur Optimierung der SVM-Parameter das Verfahren SMO vorgestellt.

- (a) Warum optimiert man hierbei nicht alle Parameter gleichzeitig, und wie viele Parameter werden statt dessen gleichzeitig optimiert und wieso?
- (b) Gehen Sie davon aus, dass  $C = 0, 1$  ist. Welchen Wert kann  $\alpha_2$  mindestens und welchen Wert maximal annehmen? Lösen Sie diese Aufgabe zeichnerisch und leiten Sie ihre Lösung daraus ab. Gehen Sie einmal von  $y_1 = y_2$  und dann von  $y_1 \neq y_2$  aus.

**Aufgabe 10.4 (2 Punkte)**

Die in der Vorlesung vorgestellten Verfahren  $SVM_{struct}$  sowie  $HMMs$  können beide für gewisse strukturelle Probleme angewandt werden. Betrachten Sie für diese Aufgabe Abb. 1. Es sind Beziehungen zwischen Nutzergruppen des sozialen Netzwerkes *facebook* dargestellt. Die Betreiber haben vier Nutzergruppen definiert, die sich durch bestimmte Beziehungen untereinander und gewissen eigene Attributen auszeichnen. Die eigenen Attribute sind in Tabelle 1 angegeben.

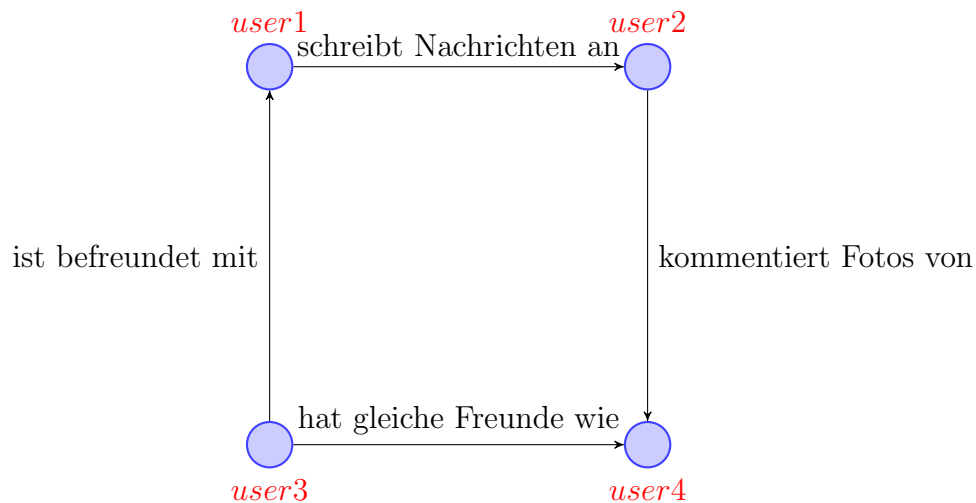


Figure 1: Beziehungen zwischen den Nutzergruppen von *facebook*

Nutzergruppe	Eigenschaften
<i>user1</i>	hat viele Freunde und kommentiert viel
<i>user2</i>	spielt Spiele und hat wenige Freunde
<i>user3</i>	kommentiert nie
<i>user4</i>	hat viele Fotos

Table 1: Eigenschaften der vier Nutzergruppen von *facebook*

- (a) Verarbeiten Sie die Beziehungen und die Eigenschaften direkt in einem  $HMM$ -Modell. Warum ist dieses (triviale)  $HMM$  nicht sinnvoll bzw. kaum nutzbar, und welche Graphen können nicht ad hoc in  $HMMs$  konvertiert werden?
- (b) Stellen Sie den aus der Vorlesung bekannten  $\Psi$ -Vektor für die  $SVM_{struct}$  auf, um alle Informationen, die die *facebook* Betreiber zusammengestellt haben, zu codieren!