

# Maschinelles Lernen

## 11. Übung

Prof. Tobias Scheffer  
Dr. Niels Landwehr  
Christoph Sawade  
Jules Rasetaharison

WS10/11

Ausgabe am: 24.01.10  
Besprechung am: 31.01.10

### Aufgabe 1 (1/3 Punkt):

Die Multiklassen-SVM kann durch folgendes Optimierungsproblem dargestellt werden:

$$\min_{\mathbf{w}} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i$$

unter den Nebenbedingungen:

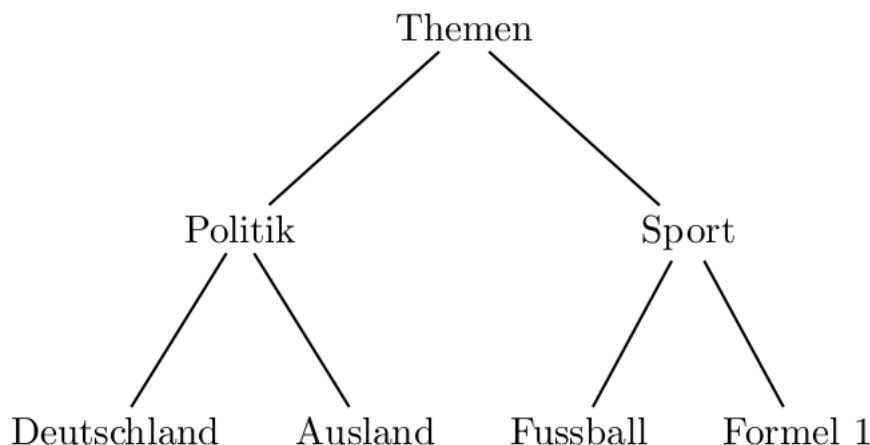
$$\begin{aligned} \forall i \in \{1, \dots, n\} \forall y \neq y_i : \langle \mathbf{w}, \Phi(\mathbf{x}_i, y_i) \rangle &\geq \langle \mathbf{w}, \Phi(\mathbf{x}_i, y) \rangle + 1 - \xi_i \\ \forall i \in \{1, \dots, n\} : \xi_i &\geq 0. \end{aligned}$$

Zeigen Sie, dass die (binäre) SVM ein Spezialfall dieser Formulierung ist. Vergleichen Sie die optimalen Gewichtsvektoren beider Lösungen.

### Aufgabe 2 (1/3 Punkt):

Geben Sie für die Probleme mit strukturierten Ausgaberräumen die gemeinsame Repräsentation von Ein- und Ausgabe  $\Phi(\mathbf{x}, \mathbf{y})$  an.

a) Sie wollen einen Klassifikator lernen, der Texte anhand ihrer Überschriften einem Thema zuordnet. Diese sind durch die folgende Hierarchie gegeben:



Geben Sie  $\Phi(x, y)$  für die folgenden Beispiele an:

- i) “Zico betroffen, Tevez trifft doppelt” → Thema: Fussball
- ii) “Merkel trifft Stoiber” → Thema: Deutschland

Welche Abschnitte des gelernten Gewichtsvektor kann man unterscheiden? Diskutieren Sie Möglichkeiten, einen Lernalgorithmus so anzupassen, dass er einen Klassifikator bevorzugt, der möglichst lange richtige Pfade vorhersagt.

b) Sie wollen einen Klassifikator lernen, der für Internetbenutzer das Alter und den Familienstand aus den Themen der von ihnen besuchten Webseiten vorhersagt. Dabei werden die diskreten Altersabschnitte  $Y_1 = \{< 12; 12 - 25; 26 - 45; > 45\}$  und Familienstände  $Y_2 = \{ledig; verheiratet; verwitwet\}$  unterschieden. Geben Sie  $\Phi(\mathbf{x}, \mathbf{y})$  für die folgenden Beispiele an:

i) *Thema: Fussball*  $\rightarrow$  Alter: 12-25, Familienstand: ledig

ii) *Thema: Deutschland*  $\rightarrow$  Alter:  $> 45$ , Familienstand: verwitwet

Welche Abschnitte des gelernten Gewichtsvektor kann man unterscheiden? Worin besteht der Vorteil dieser kollektiven Klassifikation gegenüber einer unabhängigen Vorhersage von  $Y_1$  und  $Y_2$  mit Hilfe einer Multiklassen-SVM?

**Aufgabe 3 (1/3 Punkt):**

Entwickeln sie eine Variante des Stochastische Gradient Descent-Algorithmus für das Optimierungsproblem der logistischen Regression:

$$\arg \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \log \left( 1 + \exp \left( -y_i \mathbf{w}^\top \mathbf{x}_i \right) \right).$$