

Sprachtechnologie

10. Übung

Prof. Tobias Scheffer
Thomas Vanck

Sommer 2009

Ausgabe am: 05.07.10
Besprechung am: 12.07.10

Aufgabe 1

Multiklassen-SVM

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

Multiklassen-SVM

Bisher hatten Sie in den Übungen nur Textklassifikatoren gelernt, die lediglich feindliche marsianische Texte von anderen diskriminieren können. Erich von Däniken erläutert Ihnen, dass die Marsianer auch über andere Themen diskutieren und gibt Ihnen einige Beispiele:

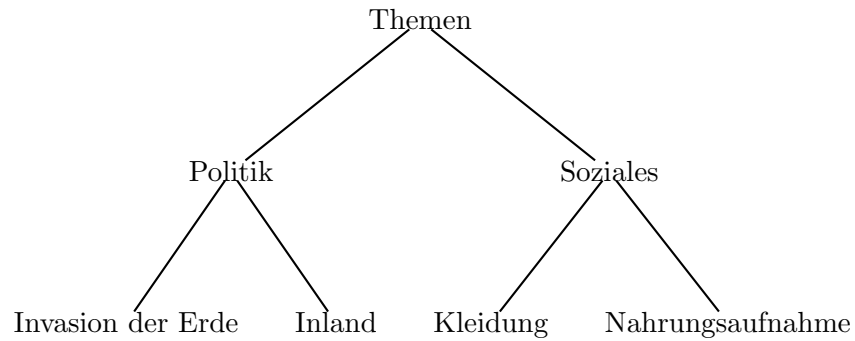
- „*argh bob argh*“, Thema: Nahrungsaufnahme
- „*zonk zonk bob*“, Thema: Invasion der Erde
- „*argh zonk bob*“, Thema: Invasion der Erde
- „*zonk zonk argh*“, Thema: Kleidung

Stellen Sie die gemeinsame Repräsentation von Ein- und Ausgabe für die Elemente der Trainingsmenge dar. Angenommen Sie würden nun eine SVM trainieren. Wie sehen die zugehörigen Nebenbedingungen aus? Welche Form hat der gelernte Klassifikator.

Aufgabe 3

Klassifikation mit Taxonomien

Erich von Däniken stellte fest, dass die Texte der Marsianer in die folgende Hierarchie eingeordnet werden können:



Angenommen Sie wollen für das Korpus aus Aufgabe 2 einen Klassifikator lernen, der die Texte in diese Baumstruktur einordnet. Wie sehen die gemeinsamen Repräsentationen von Ein- und Ausgabe für diese Beispiele aus? Welche Abschnitte des gelernten Gewichtsvektor kann man unterscheiden? Diskutieren Sie Möglichkeiten, den Lernalgorithmus so anzupassen, dass er einen Klassifikator bevorzugt, der möglichst lange richtige Pfade vorhersagt, d.h. für das erste Beispiel „Kleidung“ vorherzusagen, soll weniger bestraft werden als „Invasion der Erde“.