

# Sprachtechnologie

## 7. Übung

Prof. Tobias Scheffer  
Thomas Vanck

Sommer 2010

Ausgabe am: 14.06.10  
Besprechung am: 21.06.10

### Aufgabe 1

*Lineare Klassifikatoren*

Simulieren Sie das Training eines Rocchio- und Perzeptron-Klassifikators mit den TF-IDF-Vektoren, die Sie in Aufgabe 1 des 6. Übungsblatts berechnet haben, von Hand und stellen Sie die erhaltenen Modelle graphisch dar.

### Aufgabe 2

*Margin-Perzeptron*

Trainieren Sie für die Beispiele in Aufgabe 1 (6. Übung) ein Margin-Perzeptron mit  $\delta = 0.06$  und visualisieren Sie die Lösung. Berechnen Sie den Abstand der Trainingsbeispiele von der gelernten Hyperebene. Was ist der größtmögliche Wert  $\delta^*$  für den der Algorithmus noch terminiert? Machen Sie an einem Beispiel deutlich, welchen Vorteil das Margin-Perzeptron gegenüber den früheren Verfahren hat.

*Hinweis:* Für  $b = 0$  sind maximal sechs Iterationen notwendig.

### Aufgabe 3

*Support Vector Machine*

In der Vorlesung wurde die SVM vorgestellt. Zeigen Sie, dass für jedes  $\delta > 0$  das Optimierungsproblem

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i, \text{ mit } C > 0,$$

so dass für alle  $i \in \{1, \dots, n\}$

$$y_i \langle \mathbf{w}, \mathbf{x}_i \rangle \geq \delta - \xi_i$$

äquivalent zu der bekannten Darstellung ( $\delta = 1$ ) ist.