

Sprachtechnologie

8. Übung

Prof. Tobias Scheffer
Christoph Sawade

Sommer 2009

Ausgabe am: 23.06.09
Besprechung am: 29.06.09

Aufgabe 1

Margin-Perzeptron

Trainieren Sie für das Beispiel in Übung 6 (siehe Aufgabe 2) 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 δ^* 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 2

Primale SVM

In der Vorlesung wurde die primale 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 f.a. $i \in \{1, \dots, n\}$

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

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

Aufgabe 3

Multiklassen-SVM

Mit Ihrem bisher gelernten Textklassifikator konnten lediglich feindliche marsianische Texte von anderen diskriminiert werden. Erich von Däniken erläutert Ihnen, dass die Marsianer auch über andere Themen diskutieren und gibt Ihnen einige Beispiele:

- a) „*argh bob argh*“, Thema: Nahrungsaufnahme
- b) „*zonk zonk bob*“, Thema: Invasion der Erde
- c) „*argh zonk bob*“, Thema: Invasion der Erde
- d) „*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.