

Sprachtechnologie

10. Übung

Prof. Tobias Scheffer
Uwe Dick

Sommer 2015

Ausgabe am: 29.06.15
Besprechung am: 06.07.15

Aufgabe 1

Lineare Klassifikatoren

ID	1	2	3	4
TF-IDF	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.9 \\ 0.4 \end{pmatrix}$	$\begin{pmatrix} 0.7 \\ 0.6 \end{pmatrix}$
Klasse	+1	-1	+1	-1

Simulieren Sie das Training eines Rocchio- und eines Perzeptron-Klassifikators mit den obigen Trainingsdaten von Hand und stellen Sie die erhaltenen Modelle graphisch dar (Trennebene und Trainingsbeispiele im Koordinatensystem).

Aufgabe 2

Margin-Perzeptron

Trainieren Sie für die obige Trainingsmenge ein Margin-Perzeptron mit $\delta = 0.15$ 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.

Aufgabe 3

Multiklassen-SVM

Die gelernten Textklassifikatoren aus der letzten Übung konnten lediglich feindliche marsianische Texte von anderen unterscheiden. 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

Die Liste der relevanten Terme umfasst wieder nur *argh* und *zonk*. Stellen Sie die gemeinsame Repräsentation von Ein- und Ausgabe für die Elemente der Trainingsmenge dar, wenn Sie als Repräsentation $\phi(x)$ die (natürliche) Termfrequenz nehmen (Vektorraummodell-Folien S.11). Angenommen Sie würden nun eine SVM trainieren. Wie sehen die zugehörigen Nebenbedingungen aus? Welche Form hat der gelernte Klassifikator?