

# Maschinelles Lernen und Data Mining

## 10. Übung

Prof. Tobias Scheffer  
Steffen Bickel

WS06/07

Ausgabe am: 22.01.07  
Besprechung am: 26.01.07

### Aufgabe 1 (1/5 Punkte):

Beim Lernen über strukturierten Ausgaberräumen mit der Hidden Markov Support Vector Machine werden negative Pseudobeispiele als Ersatz für eine riesige Menge negativer Beispiele generiert. In dem Lernverfahren für die ordinale Regression aus der Vorlesung werden keine Pseudobeispiele, sondern alle negativen Beispiele direkt verwendet. Wie könnte man hier negative Pseudobeispiele erzeugen und welche Vor- oder Nachteile hätte dies?

### Aufgabe 2 (4/5 Punkte):

Wir haben vier marsianische Nachrichten aus dem Weltall empfangen. Wir möchten die Nachrichten in zwei Gruppen aufteilen, um einen besseren Überblick zu bekommen. Dazu verwenden wir den EM-Clusteralgorithmus aus der Vorlesung. Die vier Nachrichten sind  $d_1 = \{argh, bob, zonk\}$ ,  $d_2 = \{argh, zonk, pfos\}$ ,  $d_3 = \{pfos, blin\}$ ,  $d_4 = \{argh, pfos, blin\}$ . Die geratenen Anfangsparameter  $\Theta_0$  sind:

- $\theta^1 = \theta^2 = 0.5$
- $\theta^1_{argh} = \theta^1_{bob} = \theta^1_{zonk} = \theta^1_{pfos} = \theta^1_{blin} = 0.2$
- $\theta^2_{argh} = \theta^2_{bob} = \theta^2_{zonk} = 0.14$ ;  $\theta^2_{pfos} = \theta^2_{blin} = 0.29$

Die Parameter der Dirichlet-Prior sind  $\alpha_1 = \alpha_2 = 0.1$  und  $\alpha_\omega^\varphi = 0.1$  für  $\varphi \in \{1, 2\}$  und  $\omega \in \{argh, bob, zonk, pfos, blin\}$ .

- Führen sie den ersten E-Schritt basierend auf den Parametern  $\Theta_0$  durch. Berechnen sie dazu den Posterior  $P(\phi_j = \varphi | w_{j1}, \dots, w_{jk}, \Theta_t)$  für alle Dokumente und alle Cluster.
- Führen sie den ersten M-Schritt durch und errechnen sie die Parameter  $\Theta_1$  basierend auf den Ergebnissen des ersten E-Schritts.
- Wenn man diese E- und M-Schritte wiederholt, bis die Parameter konvergieren, ergeben sich folgende Parameter  $\Theta_{32}$ :

- $\theta^1 = 0.44$ ;  $\theta^2 = 0.56$
- $\theta^1_{argh} = 0.32$ ;  $\theta^1_{bob} = 0.19$ ;  $\theta^1_{zonk} = 0.31$ ;  $\theta^1_{pfos} = 0.15$ ;  $\theta^1_{blin} = 0.03$
- $\theta^2_{argh} = 0.22$ ;  $\theta^2_{bob} = 0.02$ ;  $\theta^2_{zonk} = 0.07$ ;  $\theta^2_{pfos} = 0.37$ ;  $\theta^2_{blin} = 0.32$

Die dazugehörigen Posterior sind:

- $P(\phi_1 = 1 | W_{j=1}, \Theta) = 0.98$ ;  $P(\phi_2 = 1 | W_{j=2}, \Theta) = 0.66$ ;  
 $P(\phi_3 = 1 | W_{j=3}, \Theta) = 0.03$ ;  $P(\phi_4 = 1 | W_{j=4}, \Theta) = 0.04$
- $P(\phi_1 = 2 | W_{j=1}, \Theta) = 0.02$ ;  $P(\phi_2 = 2 | W_{j=2}, \Theta) = 0.34$ ;  
 $P(\phi_3 = 2 | W_{j=3}, \Theta) = 0.97$ ;  $P(\phi_4 = 2 | W_{j=4}, \Theta) = 0.96$

Wie lassen sich diese Ergebnisse interpretieren, welches Dokument befindet sich in welchem Cluster? Um ein (lokales) Optimum welcher Funktion handelt es sich bei diesen Ergebnissen?

- Warum ist es wichtig, dass die geratenen Anfangsparameter für alle Cluster verschieden sind?