Maschinelles Lernen II

6. Übung

Prof. Tobias Scheffer
Dr. Niels Landwehr
Christoph Sawade

Sommer 2010

Ausgabe am: 02.06.10
Besprechung am: 09.06.10

Aufgabe 1 *Spectral Clustering*

Eigenvektoren u_i sind durch die folgende Gleichung definiert:

$$u_i^T L u_i = u_i^T \lambda_i u_i,$$

wobei λ_i die entsprechenden Eigenwerte darstellen. Beweisen Sie folgenden Satz aus der Graphentheorie:

Sei G ein ungerichteter Graph mit nicht-negativen Kantengewichten. Die Anzahl der Zusammenhangskomponenten $A_1, \ldots, A_k$ von G entspricht der Anzahl der Eigenvektoren u_i der Laplace-Matrix L mit Eigenwert Null.

Hinweis: Beweis per vollständiger Induktion. Beachten Sie, dass Permutation von Zeilen und Spalten keinen Einfluss auf die Lösung des Eigenwertproblems haben.

Aufgabe 2 *Hierarchisches Clustering*

Gegeben seien folgende Instanzen $V = \{\mathbf{x}_1, \ldots, \mathbf{x}_9\}$.

i	1	2	3	4	5	6	7	8	9
$[\mathbf{x}_i]_1$	$0,0$	$0,1$	$0,2$	$0,4$	$0,6$	$0,8$	$1,2$	$1,6$	$2,0$

- Wenden Sie den Algorithmus AGNES auf V an und geben Sie das Resultat für die erwartete Clusteranzahl $k = 3, \ldots, 9$ an.
- Wenden Sie den Algorithmus DIANA auf V an und geben Sie das Resultat für die erwartete Clusteranzahl $k = 1, 2, 3$ an.

Aufgabe 3 *Evaluierung von Clusterings*

Gegeben sei folgendes Clustering $\mathcal{C}^* = \{\{\mathbf{x}_1, \ldots, \mathbf{x}_3\}, \{\mathbf{x}_4, \ldots, \mathbf{x}_6\}, \{\mathbf{x}_7, \ldots, \mathbf{x}_9\}\}$. Berechnen Sie für $\mathcal{C}' = \{\{\mathbf{x}_1, \ldots, \mathbf{x}_6\}, \{\mathbf{x}_7, \mathbf{x}_8\}, \{\mathbf{x}_9\}\}$

- die relative Entropie,
- das durchschnittliche F-measure und
- den Rand-Index

bzgl. der Evaluationsdaten $\mathcal{C}^*$.