

Maschinelles Lernen und Data Mining

12. Übung

WS06/07

Prof. Tobias Scheffer
Steffen Bickel

Ausgabe am: 02.02.07
Besprechung am: 09.02.07

Aufgabe 1 (3/3 Punkt):

Nachdem sie nun viele verschiedene maschinelle Lernverfahren in der Vorlesung kennengelernt haben, möchten sie nun endlich eine Millionen Dollar in dem aktuell laufenden Netflix-Wettbewerb gewinnen (www.netflixprize.com). Das Ziel des Wettbewerbs ist, Filmbewertungen verschiedener Netflix-Kunden möglichst gut vorherzusagen.

- (a) Welche Problemstellungen und Algorithmen aus der Vorlesung haben einen Bezug zu der Problemstellung oder Teilproblemen des Netflix-Wettbewerbs? Diskutieren sie die Übereinstimmungen und Unterschiede der Netflix-Problemstellung zu den Problemstellungen aus der Vorlesung.
- (b) Überlegen sie sich nun eine Lösungsstrategie für den Wettbewerb. Beschreiben sie möglichst genau, welchen Algorithmus sie wie einsetzen möchten. Beachten sie auch, dass ihre Lösung in der Anzahl der Trainingsdaten gut skalieren sollte, denn es werden 100 Millionen Ratings vorgegeben.
- (c) Wie könnte man externe Daten, z.B. die IMDB-Movie Database (www.imdb.com), nutzen, um ihre Lösung aus b) zu verbessern?