

Anna Östlin Pagh and Rasmus Pagh
IT University of Copenhagen

Advanced Database Technology

April 18, 2005

DATA MINING

Lecture based on [GUW 20.6], [BrinPage98, sec. 1, 2, 4.2],
[MethaAgrawalRissanen96], [AgrawalSrikant94], and [Ullman00]

Today

- Advertisements.
- Data mining
 - Introduction.
 - Data mining in Google.
 - Association rules: A priori algorithm and improvements.
 - Classification: The SLIQ classifier.

Advertisements

- Two courses to watch out for in fall 2005:
 - "Advanced algorithms".
 - Course on "hard" AI (don't know the title).
- Teaching assistants needed for "Introduction to databases", fall 2005. If you are interested, send me an e-mail.

Data mining: Many flavors

- Statistics.
- Artificial intelligence.
- OLAP/dimensional modeling: Complex aggregation queries over possibly huge data sets ("decision support").
- Here: "Discovery of useful summaries of data".

Examples of data mining queries

- **Clustering.** Group objects together in clusters of "similar" objects, e.g., customer groups that need different treatment.
- **Association rules.** Find "interesting" correlations in data. Amazon.com: "Customers who bought this book also bought ..."

Data mining in this lecture

- Google's method for ranking web pages according to how authoritative they are.
- Finding association rules
 - The A-Priori algorithm...
 - and improvements
- Classification:
 - Decision tree classification
 - The SLIQ classifier