Datamining Applications

IAP Course
CBCL

Statistical Learning Becomes Hot

- Amount of data grows (terabytes/day!)
- How to: find patterns, make predictions, develop models, search, filter…. ?
- Many problems are naturally set up as "learning problems"

Information Retrieval

- Search in image, video, music… databases
- Internet search engines
- Scientific data retrieval
- Personal data management

Information Filtering

- Classify web pages and control permissions
- Email categorization for appropriate routing or automatic response
- News filtering

Fraud Detection

Case Study: detect credit card fraud

- INPUT
- Past actions
- Behavior of past fraud
- Background of user
- …..

- OUTPUT
- A credit card fraud
- An auction risk
- A mobile phone fraud

Web-Clicks Analysis

- Differentiate transactors from visitors
- Differentiate transaction types (is it a valid transaction?)
- Personalization of content of pages
- Personalization of buying opportunities

Datamining in 1984

- Customer profiling
- Direct marketing
- Selection of job applicants
- Insurance: classify customers
- Surveillance (video, email….)
- ……..

Other applications

- Churning: the process whereby customers switch providers
- Predict audience share for TV
- Predict sales for inventory control

Applications at CBCL

- Image analysis: e.g. face detection in image
- Text categorization: e.g. classify news
- Multimedia analysis (video, sound, …)
- Bioinformatics: study diseases, genome
- Time series (e.g. financial data) analysis