Lecture 4: Knowledge Management Tools

Dr. Roman V Belavkin

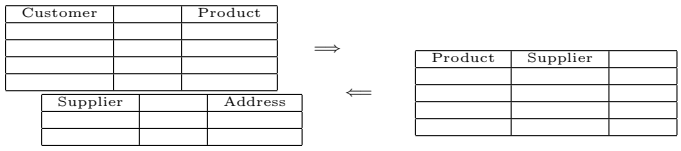
BIS4410

Contents

1 Knowledge Creation and Capture

Operational Databases

Most operational DBs rely on the *relational* model (two-dimensional tables):



These are not designed for complex queries required by managers.

Table 1: GBP/EUR rates 4–8 Jan, 2010

Date	Today	Tomorrow
2010/01/04	0.89513	0.89966
2010/01/05	0.89966	0.89934
2010/01/06	0.89934	0.89963
2010/01/07	0.89963	0.89771
2010/01/08	0.89771	?

The Nature of Management Queries

Example 1. For the worst performing region? Is there a problem with suppliers?

- Need to use operational data. Possibly from many DBs at once and over a long period of time
- Need new and more complex queries
- Rapid results
- Should be able to analyse and follow up the results
- Must not disrupt the operational running

Example: Credit Score

Case:	Age	Gender	M. Income (£ K)	M. Expenses (£ K)	Home owner	Credit score
1	21	0	2	1	0	3
2	18	1	1	2	0	1
3	50	1	6	2	1	5
4	23	0	3	1	1	4
5	40	1	3	2	0	2

- Does the credit score depend on a person's income?
- Can we find a function $f(\cdot)$ such that

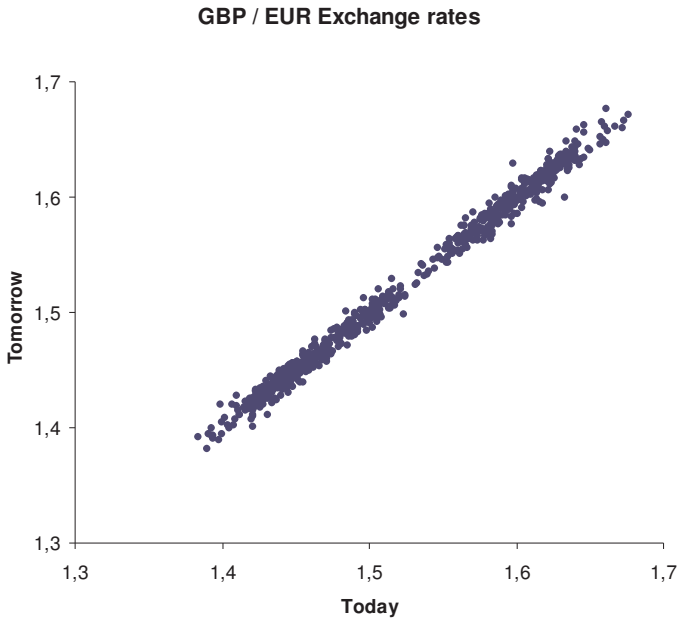
$$\text{Credit score} = f(\text{Income, Expenses, Age, Gender}, \dots)$$

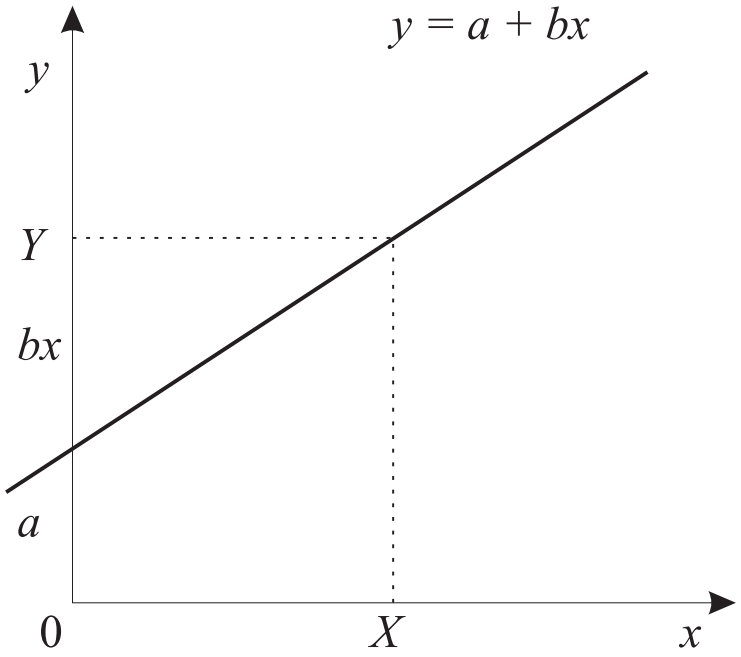
- Data-driven modelling is a search for such functions that represent the dependencies between different variables.

Example: Exchange Rates

If there are just two variables x (e.g. 'Today') and y (e.g. 'Tomorrow'), then we can use a function $f(x)$ of one variable to model y:

$$\text{Tomorrow} \approx f(\text{Today})$$



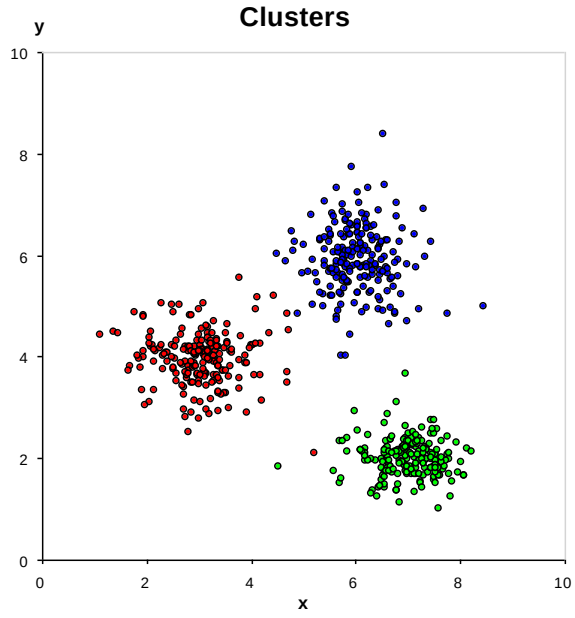


For example, we can use linear model with parameters a (intercept) and b (slope):

$$y \approx f(x) = a + b\,x$$

Example: Clustering

Case:	Age	Gender	M. Income (£ K)	M. Expenses (£ K)	Home owner	Credit score
1	21	0	2	1	0	3
2	18	1	1	2	0	1
3	50	1	6	2	1	5
4	23	0	3	1	1	4
5	40	1	3	2	0	2

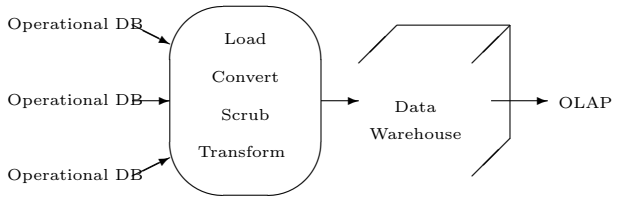


- **Clusters** are groups of points close to each other.
- One of the main goals of **multivariate analysis** is to find clusters of points.
- 'Similar' customers would have small distance between them and would belong to the same group (cluster).

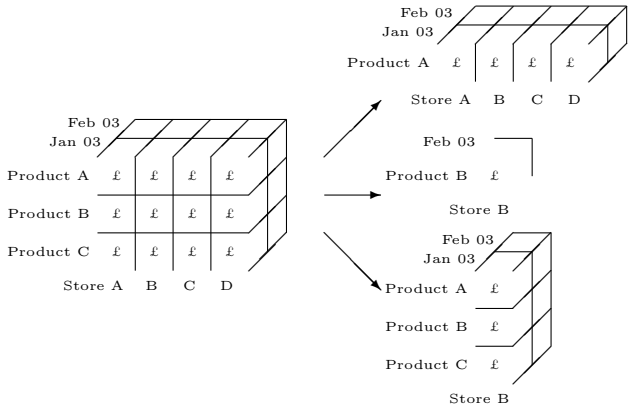
Data Warehouse

- *Data warehouse* (DW) is a centralised repository of all corporate data
- The purpose of DW is to provide managers with information to support business decisions (not business operations!)
- The long-term storage of data enables to see trends and make forecasts

Data Warehouse Structure



Data Cube



Textual Data

- Digital libraries, journals
- Online bulletin and discussion boards
- Mailing lists
- Blogs
- Newsfeeds

2 Knowledge Acquisition and Application Phase

Decision Support Systems

Definition 2 (Decision Support Systems (DSS))**.** Information systems designed to support mangers and organisations in making-decisions

- Typical components:
  1. Database or knowledge base
  2. Model
  3. User interface
- Can be classified by levels: Operational, tactical, strategic (executive)

Expert Systems

Definition 3 (Expert Systems (ES))**.** Computer programs that try to replicate knowledge and skills of human experts in some area, and then solve problems in this area (the way human experts would).

- ES take their roots in *Cognitive Science* — the study of human mind using combination of AI and psychology.
- ES were the first successful applications of AI to real-world problems solving problems in medicine, chemistry, finance and even in space (Space Shuttle, robots on other planets).
- ES contribute to savings of $ millions
- There are *rule-based* and *case-based* reasoning systems

Data-driven models and simulations

- The word '*model*' comes from a Latin word meaning '*small*', and usually we mean some small representation of the real object (e.g. a model of a house, a car or an aircraft)

$$\text{Reality} > \text{Model}$$

- Statistical modelling tools (e.g. SPSS, R, Matlab)
- Monte-Carlo simulations of stochastic processes (e.g. GPSS)

OLAP

- *Online Analytical Processing* (OLAP) is a system that further transforms the data into a more structured form than tables.
- OLAP is a form of EIS. Many managers can access a data warehouse simultaneously through OLAP (e.g. by Internet).
- The data in OLAP is usually stored in a form of multidimensional cubes (hypercubes).
- Many calculations have been already performed and stored in the hypercube (e.g. totals, aggregate data, etc)
- OLAPs provide managers with a rapid and flexible access to data without them having to be programmers

3 Knowledge Sharing and Dissemination Phase

Groupware and Collaboration Tools

Definition 4 (Groupware)**.** Hardware and software connected to a network to support activities of a group of colleagues.

- Scheduling and planning
- E-mail, telephone utilities
- Newsletters and mailing lists
- File repositories (e.g. DropBox)
- Teleconferences

Classification of Groupware Technologies

	Same TimeSynchronous	Different TimeAsynchronous
Same place co-located	Voting, Presentation support	Shared computers
Different place-distant	Videophones, Chat	E-mail, Workflow

Network Technologies

- Knowledge repositories and knowledge portals: External, structured internal, informal internal
- Wikis
- May contain declarative (what), procedural (how), causal (why), context (why care)

Reading

Additional reading for the following week is Grobelnik and Mladenic (2005).

References

Grobelnik, M., & Mladenic, D. (2005). Automated knowledge discovery in advanced knowledge management. *Journal of knowledge management*, *9*, 132–141.