# Data Science

# Course Info

Themis Palpanas

University of Paris

*http://www.mi.parisdescartes.fr/~themisp/*

1

# Course Prerequisites

- Databases or Data Management

- basic probability and statistics

2

1

# Course Material

- course taught using:
  - slides
  - published research papers

- course-related material and information:
  - disseminated through class moodle page

# Course Requirements

- Class Participation

- Quizzes

- Project

- Final Exam

# Course Requirements

- Class Participation
  - participate in class discussions of problems and research papers
- Quizzes
- Project
- Final Exam

# Course Requirements

- Class Participation
- Quizzes
  - Question answering
  - Programming
- Project
- Final Exam

# Course Requirements

- Class Participation

- Quizzes

- Project

  - work on a major project to be submitted at the end of the course (with intermediate deadlines)

    - detailed marking scheme on course web page

- Final Exam

# Course Requirements

- Class Participation

- Quizzes

- Project

- Final Exam

  - oral examination on the submitted project

# Project

- Software implementation related to course material

- Should involve an original component or experiment

- More later about available data and computing resources.

- Groups of 3 persons

- project must be completed by the end of the course

10

# Recommended Reference Books

- S. Chakrabarti. Mining the Web: Statistical Analysis of Hypertex and Semi-Structured Data. Morgan Kaufmann, 2002
- R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, 2ed., Wiley-Interscience, 2000
- T. Dasu and T. Johnson. Exploratory Data Mining and Data Cleaning. John Wiley & Sons, 2003
- U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press, 1996
- U. Fayyad, G. Grinstein, and A. Wierse, Information Visualization in Data Mining and Knowledge Discovery, Morgan Kaufmann, 2001
- **J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2nd ed., 2006**
- D. J. Hand, H. Mannila, and P. Smyth, Principles of Data Mining, MIT Press, 2001
- **T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer-Verlag, 2001**
- T. M. Mitchell, Machine Learning, McGraw Hill, 1997
- G. Piatetsky-Shapiro and W. J. Frawley. Knowledge Discovery in Databases. AAAI/MIT Press, 1991
- **P.-N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining, Wiley, 2005**
- S. M. Weiss and N. Indurkhya, Predictive Data Mining, Morgan Kaufmann, 1998
- **I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, 2nd ed. 2005**

11