



Data Mining Course

Master in Information Technologies

Enginyeria Informàtica

Facultat d'Informàtica de Barcelona

L. Belanche⁽²⁾ , K. Gibert⁽¹⁾

⁽¹⁾Department of Software, Soft-computing group

⁽²⁾Department of Statistics and Operation Research

Knowledge Engineering and Machine Learning group

Universitat Politècnica de Catalunya, Barcelona

K. Gibert



Practical information

- Lecturers
 - K. Gibert (C5, 2nd Floor, karina.gibert@upc.edu)
 - L. Belanche
- Materials
 - T. Aluja, L. Belanche, K. Gibert
- Structure
 - Theory sessions: Mon (5-6 pm) Wed (5-7 pm) English
 - Practical sessions:
 - Mon (6-8 pm) Catalan
 - Fry (3-5 pm) English
 - Guided practice + practical work assessment

Topics

- Introduction
- Preprocessing
- Profiling
- Visualization
- Clustering
- Association Rules
- Decision Trees
- Parametric models
- Non-parametric models
- Neural Networks
- Support Vector Machines



See Schedule in the web

Evaluation

Final grade = f(Theoretical test, practical works)

Theoretical test: 8 short questions about the course topics

- T1: 4 from the first part (K. Gibert)
- T2: 4 from the second part (L. Belanche)

Individual prove

Practical works: Three practical works developed by pairs of students

- P1: Data cleaning and preprocessing (K. Gibert)
- P2: Data Modelling (non-supervised) (K. Gibert)

Single written report for P1+P2

12 nov 2012

- P3: Complete Data Mining process (supervised) (L. Belanche)

Written report for P3 + oral presentation

14 jan 2013

21 jan 2012

$$\begin{aligned}\text{Final grade} &= 0.1 T1 + 0.1 T2 + 0.15 P1 + 0.15 P2 + 0.5 P3 = \\ &= 0.1 T1 + 0.1 T2 + 0.3 (P1 + P2) + 0.5 P3\end{aligned}$$

Materials

- See link in the raco:
 - Slides, datasets, practical works guidelines...
- References

Aluja T., Morineau A. **Aprender de los datos: El Análisis de Componentes Principales**, EUB, 1999.

Hand D.J. **Construction and Assessment of Classification Rules.** , John Wiley, 1997.

Hastie T., Tibshirani R., Friedman J. **The elements of statistical learning. Data mining, inference and prediction.** , Springer, 2001.

Hernández Orallo J., Ramírez Quintana M.J., Ferri Ramírez C **Introducción a la Minería de Datos**, Prentice Hall, 2004.

Witten I.H., Frank E **Data Mining**, . Morgan Kaufman Publishers, 2000.

Berry M.J.A., Linoff G **Data Mining Techniques, for marketing, sales and costumer support**, John Wiley, 1997.

Hand D., Mannila H., Smyth P. **Principles of Data Mining**, The MIT Press, 2001.

Ripley B.D. **Pattern Recognition and Neural Networks.** , Cambridge University Press, 1995.

Bishop C. M. **Neural Networks for Pattern Recognition**, Clarendon Press. Oxford, 1995.

Cyos, K., Pedyioz, W. I Swiniaski, R. **Data Mining. Methods for Knowledge Discovery**, Kluwer, 1998.

Course DM: Introduction. T. Aluja

3

Software

<http://www.cran.r-project.org>

<http://www.kdnuggets.com/>

<http://www.cs.waikato.ac.nz/>

<http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html>

<http://ses.telecom-paristech.fr/lebart/>

http://en.wikipedia.org/wiki/Data_mining

<http://www.itl.nist.gov/div898/handbook/pmd/pmd.htm>