

# Third CEU Summerschool on Advanced Statistics and Data Mining (June 30th–July 11th, 2008)

---

*Source:* <http://coding.derkeiler.com/Archive/General/comp.theory/2008-04/msg00012.html>

---

- *From:* coss <[cossorzano@xxxxxxxxx](mailto:cossorzano@xxxxxxxxx)>
  - *Date:* Mon, 7 Apr 2008 08:59:11 -0700 (PDT)
- 

Dear colleagues,

San Pablo – CEU University in collaboration with other five universities (Málaga, Politécnica de Madrid, País Vasco, Complutense, and Castilla La Mancha), Unión Fenosa, CSIC and IEEE organizes a summerschool on "Advanced Statistics and Data Mining" in Madrid between June 30th and July 11th. The summerschool comprises 12 courses divided in 2 weeks.

Attendees may register in each course independently. Registration will be considered upon strict arrival order. For more information, please, visit <http://biocomp.cnb.csic.es/~coss/Docencia/ADAM/ADAM.htm>.

Best regards, Carlos Oscar

\*List of courses and brief description\* (full description at <http://biocomp.cnb.csic.es/~coss/Docencia/ADAM/ADAM.htm>)

Week 1 (June 30th – July 4th, 2008)

Course 1: Bayesian networks (15 h), Practical sessions: Hugin, Elvira, Weka, LibB

Bayesian networks basics. Inference in Bayesian networks.  
Learning Bayesian networks from data

Course 2: Multivariate data analysis (15 h), Practical sessions: MATLAB

Introduction. Data Examination. Principal component analysis (PCA). Factor Analysis. Multidimensional Scaling (MDS). Correspondence analysis.

Multivariate Analysis of Variance (MANOVA). Canonical correlation.

Course 3: Supervised pattern recognition (Classification) (15 h), Practical sessions: Weka

Introduction. Assessing the Performance of Supervised Classification Algorithms.

Classification techniques. Combining Classifiers.  
Comparing Supervised Classification Algorithms

## Third CEU Summerschool on Advanced Statistics and Data Mining (June 30th–July 11th, 2008)

Course 4: Association rules (15 h), Practical sessions: Bioinformatic tools

Introduction. Association rule discovering. Rule Induction. KDD in biological data.

Applications. Hands-on exercises.

Course 5: Neural networks (15 h), Practical sessions: MATLAB

Introduction to the biological models. Nomenclature. Perceptron networks.

The Hebb rule. Foundations of multivariate optimization. Numerical optimization.

Rule of Widrow–Hoff. Backpropagation algorithm.

Practical data modelling with neural networks

Course 6: Time series analysis (15 h), Practical sessions: MATLAB

Introduction. Probability models to time series. Regression and Fourier analysis.

Forecasting and Data mining.

Week 2 (July 7th – July 11th, 2008)

Course 7: Regression (15 h), Practical sessions: SPSS

Introduction. Simple Linear Regression Model. Measures of model adequacy.

Multiple Linear Regression. Regression Diagnostics and model violations.

Polynomial regression. Variable selection. Indicator variables as regressors.

Logistic regression. Nonlinear Regression.

Course 8: Practical Statistical Questions (15 h), Practical sessions: study of cases (without computer)

I would like to know the intuitive definition and use of ...: The basics.

How do I collect the data? Experimental design.

Now I have data, how do I extract information? Parameter estimation

Can I see any interesting association between two variables, two populations, ...?

How can I know if what I see is "true"? Hypothesis testing

How many samples do I need for my test?: Sample size

Can I deduce a model for my data? Other questions?

Course 9: Hidden Markov Models (15 h), Practical sessions:HTK

Introduction. Discrete Hidden Markov Models. Basic algorithms for Hidden Markov Models.

Semicontinuous Hidden Markov Models. Continuous Hidden Markov Models.

Unit selection and clustering. Speaker and Environment Adaptation for HMMs.

Other applications of HMMs

Course 10: Statistical inference (15 h), Practical sessions: SPSS

Introduction. Some basic statistical test. Multiple testing.

Introduction to bootstrapping

Course 11: Dimensionality reduction (15 h), Practical sessions: MATLAB

Introduction. Matrix factorization methods. Clustering methods.

Projection methods.

Applications

Course 12: Unsupervised pattern recognition (clustering) (15 h),

Practical sessions: MATLAB

Introduction. Prototype–based clustering. Density–based clustering.

Graph–based clustering. Cluster evaluation. Miscellanea

.