Sciences des données – 3 : exploration et prédiction

Lecturer
Philippe GROSJEAN

Course mnemonic
BINF-Y402

ECTS credits
3 credits

Language(s) of instruction
French

Course period
First term

Campus
UMons

Course content

The pedagogical material is available online: https://wp.sciviews.org. The chapters of this AA are:

- Classification I - LDA, general principle, confusion matrice, metrics
- Classification II - corss-validation, AUC, k-nn, lvq, raport, random forest
- Classification III = svm, neural networks, initiation to deep learning
- Time series I - description, manipulation, acf, spectral analysis
- Time series II - decomposition & regularisation
- Spatial statistics, initiation, maps & krigging

Objectives (and/or specific learning outcomes)

To be able to find useful information in a large dataset using data mining and machine learning tools , to analyze correctly biological data with time-dependencies and to analyse the spatial data. To be able to present results in a reproducible way (reports) and to use professional software in data science: R, RStudio, R Markdown, git.

Pre-requisits and co-requisits

Required knowledge and skills

Bases in data science, including project management, data importation and transformation, visualization of data through graphs and writing of reproducible reports. General uni- and multivariate statistics, (generalized) linear models, nonlinear models, ACP & AFC, non supervised classification (hierarchical clustering and K-means). An update of the knowledge prior to the course can be done via the first two books of the data science courses available online at https://wp.sciviews.org.

Teaching method and learning activities

Blenbded learning. Students learn the theory at home before the exercise sessions (flipped classroom). All the exercises, at home or in class, are taken into account. In class session, the students essentially work in projects where they analyze biological data in practice, using a software environement around R and RStudio.

References, bibliography and recommended reading

Barnier, J., 2018. Introduction à R et au tidyverse (https://juba.github.io/tidyverse/index.html). Ismay, Ch. & Kim A.Y, 2018. Moderndive: An introduction to statistical and data science via R (http://moderndive.com). Wickham, H. & Grolemund, G, 2017. R for data science (http://r4ds.had.co.nz). Zar, J.H., 2010. Biostatistical analysis (5th ed.). Pearson Education, London. 944pp. Dagnelie, P., 2007. Statistique théorique et appliquée, Volumes I et II (2ème ed.). De Boeck & Larcier, Bruxelles. 511pp (vol. I) 734pp (vol. II). Venables W.N. & B.D. Ripley, 2002. Modern applied statistics with S-PLUS (4th ed.). Springer, New York, 495 pp. Legendre, P. & L. Legendre, 1998. Numerical ecology (2nd ed.). Springer Verlag, New York. 587 pp.

Course notes

Université virtuelle

Other information

Place(s) of teaching

UMons

Contact(s)

Philippe Grosjean (Philippe.Grosjean@umons.ac.be, sdd@sciviews.org), +32/065.37.34.97

Evaluation method(s)

Group work, Personal work, Practice work and Other

Evaluation method(s) (additional information)

Grading is established via ongoing assessment all along the Q1. Given that the grade is established through ongoing assessment of works that cannot be organized during the summer, there is no second session.

Determination of the mark (including the weighting of partial marks)

The different exercises and projects are used to calculate the grade. The exercises are polled together into four increasing levels of difficulty from 1 to 4. The grade must be at least 50% for exercises level 4 on one hand, and for all the exercices levels 1 to 3 on the other hand, or only the weakest grade og the two is used for this AA. Penalties are applied if more than 1/5 of the exercices are not done for each module. Given the way grading is done the presence to all sessions is mandatory. Any unjustified absence to a session will result in a 0/20 for the corresponding content.

See the course summary for details on the grade calculation by type of exercise.

Main language(s) of evaluation

French

Programmes

Programmes proposing this course at the faculty of Sciences

MA-BMOL | Master in Biochemistry and Molecular and Cell Biology | finalité Research (Charleroi)/unit 1