

## Sciences des données – 3 : exploration et prédiction

### Titulaire

Philippe GROSJEAN

### Mnémonique du cours

BINF-Y402

### Crédits ECTS

3 crédits

### Langue(s) d'enseignement

Français

### Période du cours

Premier quadrimestre

### Campus

UMons

## Contenu du cours

Le cours est entièrement disponible en ligne : <https://wp.sciviews.org>. Les chapitres de cette UE sont:

- > Classification supervisée I, LDA, principe général, matrice de confusion, métriques
- > Classification supervisée II, validation croisée, AUC, k-nn, l<sub>vq</sub>, rpart, random forest
- > Classification supervisée III, svm, réseaux de neurones, initiation au deep learning
- > Séries temporelles I, caractéristiques, manipulation, acf, analyse spectrale
- > Séries temporelles II, décomposition et régularisation
- > Statistiques spatiales, initiation, représentations cartographiques, krigeage

## Objectifs (et/ou acquis d'apprentissages spécifiques)

Etre capable de retrouver l'information pertinente dans un gros jeu de données à l'aide de techniques de fouille de données et de classification supervisée, d'analyser correctement des données biologiques ayant une dépendance temporelle et d'analyser des données spatiales. Etre capable de présenter ses résultats de manière reproductible (rédaction de rapports) et utiliser des logiciels professionnels dans le domaine de la science des données : R, RStudio, R Markdown/Quarto, git/GitHub.

## Pré-requis et co-requis

### Connaissances et compétences pré-requises

Bases en science des données, en particulier, la gestion de projets d'analyse des données, l'importation et le remaniement des données, la visualisation à l'aide de graphiques et la

rédaction de rapports reproductibles. Statistiques générales uni- et multivariées, modèles linéaires (généralisés) et non linéaires, technique d'ordination (ACP, AFC) et de classification non supervisée (CAH, K-means). Une mise à jour des compétences est possible via les deux premiers ouvrages des cours de science des données disponibles en ligne à partir de <https://wp.sciviews.org>.

## Méthodes d'enseignement et activités d'apprentissages

Enseignement hybride. Les étudiants apprennent la matière chez eux avant les séances d'exercices (classe inversée). L'ensemble des exercices tant à domicile qu'en séance sont pris en compte. En séance, les étudiants travaillent essentiellement dans des projets où ils analysent en pratique des données biologiques avec l'environnement logiciel utilisé dans le cours (R + RStudio).

## Contribution au profil d'enseignement

- > Avoir acquis, dans le domaine des sciences biologiques, et particulièrement dans le domaine de la biochimie, biologie moléculaire et cellulaire, des connaissances hautement spécialisées et intégrées ainsi que de larges compétences, prolongeant celles qui relèvent du niveau de bachelier en sciences biologiques
- > Etre capable de mener des travaux de recherche et de développement d'envergure en lien avec les sciences biologiques, en biochimie, biologie moléculaire et cellulaire;

+ Etre capable d'appliquer, de mobiliser, d'articuler et de valoriser les connaissances et les compétences acquises en vue de contribuer à la conduite et à la réalisation d'un projet

+ Faire preuve d'initiative et être capable de travailler seul et en équipe

- Gérer et mener un travail de recherche, de développement ou d'innovation

+ Etre capable d'appréhender une problématique inédite relevant des sciences biologiques en biochimie, biologie moléculaire et cellulaire et de ses applications

- Maîtriser les techniques de communication

+ Pouvoir communiquer de façon claire, structurée et argumentée, tant à l'oral qu'à l'écrit, ses conclusions, ses propositions originales ainsi que les connaissances et principes sous-jacents

+ Maîtriser les techniques de communications scientifiques écrites et orales tant en français qu'en anglais

- Développer et intégrer un fort degré d'autonomie

+ Poursuivre sa formation et développer de nouvelles compétences de façon autonome

- Appliquer une méthodologie scientifique

+ Faire preuve de rigueur, d'autonomie, de créativité, d'honnêteté intellectuelle, de sens éthique et déontologiques

## Références, bibliographie et lectures recommandées

Barnier, J., 2018. Introduction à R et au tidyverse (<https://juba.github.io/tidyverse/index.html>). Ismay, Ch. & Kim A.Y., 2018. Moderndiver: An introduction to statistical and data science via R (<http://moderndiver.com>). Wickham, H. & Golemund, 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.

## Support(s) de cours

Université virtuelle

## Autres renseignements

### Lieu(x) d'enseignement

UMons

### Contact(s)

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

## Méthode(s) d'évaluation

Travail de groupe, Travail personnel, Travail pratique, Autre et Projet

## Méthode(s) d'évaluation (complément)

L'évaluation se fait en continu tout au long du Q1. Les différents exercices et projets participent à l'établissement de la note. Etant donné que la note est basée sur une évaluation continue de travaux qu'il n'est pas possible d'organiser durant l'été, il n'y a pas de seconde session.

## Construction de la note (en ce compris, la pondération des notes partielles)

Les exercices sont regroupés par niveau de difficulté croissant de 1 à 4, avec des pondérations également plus importantes pour les exercices de niveau 4 et les interrogations en cours d'année.

Des pénalités sont appliquées si plus de 1/5 des exercices ne sont pas réalisés dans chaque module. Etant donné le type d'évaluation, la présence à l'ensemble des séances est obligatoire. Toute absence injustifiée à une séance sera sanctionnée d'un 0/20 pour le contenu concerné.

Voir le plan du cours pour le détail du calcul par type d'exercice.

## Langue(s) d'évaluation principale(s)

Français

## Programmes

### Programmes proposant ce cours à la faculté des Sciences

MA-BMOL | Master en biochimie et biologie moléculaire et cellulaire | finalité Approfondie (Site de Charleroi)/bloc 1