

A Journey Through Data Science

Intensive Short-Courses

Instructors:

Dr. Johan Van Horebeek, Jean Bernard Hayet, TBD

e-mail: {horebeek,jbhayet}@cimat.mx

Office: K225 - I101

Hours:

45 hours

Description:

Our journey will lead us to explore four topics of paramount importance for data scientists: data preparation and exploration, data visualization, prediction and regression modeling, and feed forward neural networks.

All along the course, a particular emphasis will be given to developing and discussing some of the mathematical and algorithmic foundations of the concepts and techniques presented in the course and to their application in real-life problems through case studies.

Course Goals:

On completion of the course, students will

- understand the basics of Data Science and its algorithmic foundations.
- develop skills in the application in real-life problems of the course tools

Overall requirements:

- Your intended major should include components involving Mathematics, Statistics, Data Science, or Computer Science.
- Basic experience with programming. R or Python is recommended but not required.

A Journey Through Data Science

Intensive Short-Courses

Instructors:

Dr. Johan Van Horebeek, Jean Bernard Hayet, TBD

e-mail: {horebeek,jbhayet}@cimat.mx

Office: K225 - I101

Hours:

45 hours

Course Content:

- Data exploration and data visualization
 - Review of the R language / Python
 - Data and uncertainty
 - Data wrangling
 - Visual encoding and effective graph building
 - Grammar of graphics
 - Visualizing 1D data
 - Visualizing multidimensional data
 - Introduction to interactivity and storytelling
- Prediction and regression modeling
 - Introduction to statistical modeling and prediction
 - Linear models
 - Nearest neighbor models for classification and regression
 - Bias-variance tradeoff
- Neural networks
 - Feed forward neural networks
 - The math behind back-propagation
 - Deep neural networks
 - Optimization techniques
 - Practical aspects and examples
 - Some recent ideas

A Journey Through Data Science

Intensive Short-Courses

Instructors:

Dr. Johan Van Horebeek, Jean Bernard Hayet, TBD

e-mail: {horebeek,jbhayet}@cimat.mx

Office: K225 - I101

Hours:

45 hours

Grading:

- Homework (40%)
- An extended written report and oral presentation of an integrative project (60%)

Bibliography:

- Wickham, H., Grolemund, H. W. G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize and Model Data*. O'Reilly
- Healy, K, (2018). *Data Visualization*. Princeton University Press
- James, G., Hastie, T., Tibshirani, R., Witten, D. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer
- James, G., Hastie, T., Tibshirani, R., Witten, D. , Taylor, J. (2023). *An Introduction to Statistical Learning: with Applications in Python*. Springer