Data analysis

Spatio-temporal data and hierarchical models.

Syllabus

Djemel Ziou

Motivations

- Nowadays, it is easy to produce data(text, document, audio, image, video).
- The amount of existing data is huge.
- 30% the increase rate/year.
- Data are an invaluable source to better understand the different scientific, economic, social and cultural phenomena.
- We need to acquire, save, transmit, analyse, summarize … data.
- Depending on the methodology and uses, we talk about scientific data analysis, business intelligence, data warehouse, information retrieval, epidemiology, data scientist, machine learning...
- Courses are required to reinforce existing training in data analysis.

Objective

- Mastering the models of data analysis indispensable for the study of spatial, temporal and spatiotemporal phenomena.
- The approach consists in inferring knowledge from uncertain measurements of the properties of these phenomena.
- The uncertainty concerns the data, the underlying processes and the parameters of the models used.
- The probabilistic approach will be used to move from measurement to knowledge.
- The student will learn to construct probabilistic hierarchical models of data, random processes and parameters. She / he will also learn to implement these models in the case of real applications.

Content

- Bayesian reasoning, hierarchical random processes, time series, kriging, regression, estimators, machine learning, feature selection and mixtures of laws of probabilities.
- The concepts will be illustrated by applications in various areas of telephony, social networks, health, imaging and geography.
- Either R or Python Languages.
- References are given for each chapter.

Goal

- Prerequisite knowledge
  - Descriptive statistics
  - Calculus
  - Programming

- Expectations
  - Go further than black box based methods
  - Simple model writing
  - Understanding papers

Knowledge Assessment System

- Homework (25%)
  - Several, but only three will be considered.
  - Groups of three students.
  - Exams will include related questions to homework.
  - Deliverable: Code sources (if any) and a report.
  - No extension allowed without an official justification.
- Project (35%)
  - Each student is free to choose the subject of his project (e.g. in connection with his research work).
  - Groups of three students.
  - I will choose randomly students and ask them questions about the project. In this case, the score of the group will be revisited.
  - Deliverable: Code sources and a report.
  - No extension allowed without an official justification.

Knowledge Assessment System

- Mid-term exam (20%) and a final exam (20%).

- Important
  - You need to produce a personal work.
  - For the plagiarism, please read the SIAT rules and the web site plagiarism.org

Teaching approach

- Research oriented course.
- PowerPoint presentation to expose the material.
- Three hours/week (sometimes more).
- At least 2 hours of work for each hour in class.
- For any concern that you have, please come and talk with me. The office hours will be announced the next class.
- I will not answer technical questions by email, WeChat…
- You Cell phone should be closed.