



E0 259 Aug 3:1

Data Analytics

Instructor

Ramesh Hariharan and Rajesh Sundaresan

Email: rajeshs@iisc.ac.in

Teaching Assistant

Nihesh Rathod Sumanbhai, A Y Sarath

Email: niheshkumar@iisc.ac.in, sarathay@iisc.ac.in

Department: ECE and CSA

Course Time: Tue, Thu., 9:30 - 11:00 AM

Lecture venue: CSA 117

Detailed Course Page: <http://www.ece.iisc.ernet.in/~rajeshs/E0259.htm>

Announcements

Brief description of the course

Data Analytics has assumed increasing importance in recent times. Several industries are now built around the use of data for decision making. Several research areas too, genomics and neuroscience being notable examples, are increasingly focused on large-scale data generation rather than small-scale experimentation to generate initial hypotheses. This brings about a need for data analytics. This course will develop modern statistical tools and modeling techniques through hands-on data analysis in a variety of application domains.

The course will illustrate the principles of hands-on data analytics through several case studies (10 such studies). On each topic, we will introduce a scientific question and discuss why it should be addressed. Next, we will present the available data, how it was collected, etc. We will then discuss models, provide analyses, and finally touch upon how to address the scientific question using the analyses.

We will cover the following case studies.

Astronomy: From Tycho Brahe's observations to the conclusion that Mars moves in an elliptical orbit.

Visual Neuroscience: Neural correlates predict search difficulty.

Genomics: Understanding the causes of cancer.

Sports: The Duckworth-Lewis-Stern method for setting targets in shortened limited overs cricket matches.

Genomics: The basis for red-green colour blindness.

Genomics: Population history of India.

Signal Processing: Video background separation.

Networks: Community detection.

Recommendation systems.

Prerequisites

Random Processes (E2 202) OR Probability and Statistics (E0 232) OR equivalent.

Syllabus

Data sets from astronomy, genomics, neuroscience, sports, surveillance cameras, and social networks will be analysed to answer specific scientific questions. Statistical tools and modeling techniques will be introduced as needed to analyse the data and eventually address the scientific question.

Course outcomes

Students will learn modeling techniques, key statistical principles, data handling techniques, will get hands-on experience with large data sets, and will learn to program in a language like Python.

Grading policy

50% for assignments, 30% for project and presentation, 20% for final examination.

Assignments

Mars orbit, visual neuroscience and oddball detection, retinoblastoma and age-related cancer, colour blindness, the Duckworth-Lewis-Stern method, community detection, population history of India.

Resources

<http://www.ece.iisc.ernet.in/~rajeshs/E0259.htm>