Indian Institute of Science

Department of Bioengineering

BE, Third Floor, Biological Sciences Building, Indian Institute of Science, Bengaluru, 560012, India
+91 80 2293 2624 chair.be@iisc.ac.in http://www.be.iisc.ac.in

Course Title: Introduction to Data Sciences for Bioengineers
Course Code: BE 221
Couse Schedule: January Semester
Credits: 3:0

Course Instructor: Siddharth Jhunjhunwala and Narendra Dixit

Pre-Requisites:

1. Undergraduate level course in probability and statistics or biostatistics; OR
2. BE-207 (Mathematical Methods for Bioengineers); OR
3. DB-201 (Mathematics and Statistics for Biologists)

Description

Bioengineering research often generates large amounts of data, analysis of which requires sound technical knowledge of data sciences. The goal of this course is to introduce students to the basic concepts and tools of statistical and machine learning, which may be useful to analyse the data generated by the medical, biological, and bioengineering community. The following topics will be covered: introduction to descriptive statistics, introduction to probability theory, discrete and continuous probability distributions, estimation, hypothesis testing, introduction to statistical learning, linear regression, analysis of categorical data, logistic regression, linear-discriminant analysis and KNN method, datasets and resampling, dimensionality reduction, support vector machines, unsupervised learning including machine learning. Problems will be presented and solved using R.

References

There is no prescribed textbook for this course. But the following reference texts are suggested:

1. Fundamentals of Biostatistics, Bernard Rosner
2. An Introduction to Statistical Learning, Gareth James et al.

Course Outcomes

Upon completion of the course, students will be able to:

1. Appreciate the contributions of statistics to data analysis
2. Apply the right hypothesis testing tools to different forms of data
3. Analyse medical and biological data using supervised learning techniques
4. Utilize unsupervised learning on large data