DATA ANALYSIS FOR BIOLOGISTS

PROF. BIPLAB BOSE

Department of Biosciences and Bioengineering

IIT Guwahati

INTENDED AUDIENCE : Students of different areas of Biology, Biotechnology, and allied subjects

INDUSTRIES APPLICABLE TO : Data analysis is an essential component in any bio-pharma/healthcare industry. Data analysis in biology has already moved out of the domain of conventional statistics, and it is expected that a student of biology is acquainted with basic concepts of modern data analysis tools.

COURSE OUTLINE :

Analysis of data is an integral part of biology, both in academic research and the Industry. With the advent of high-throughput techniques, biological data analysis has crossed the realm of classical statistical techniques and now involves techniques used by the wider data analytic and machine learning community. It is now expected that every biology student is acquainted with the key concepts and tools of data analysis. This course is designed specifically for biology students to learn the key concepts, applications, and limitations of commonly used data analysis techniques. This course emphasizes visualization and analysis of higher-dimensional data, like clustering, classification, and dimensionality reduction.

ABOUT INSTRUCTOR :

Prof. Biplab Bose is an Associate Professor in the Department of Biosciences and Bioengineering at IIT Guwahati. He has developed and taught courses on data analysis, systems biology, and bioinformatics. He is interested in understating the design principles of molecular networks, applications of dynamical systems theory and statistical physics in biology. He has also developed software like FlowPy, CorNetMap, and DEBay.

COURSE PLAN :

 Week 1: Basic concepts of probability and statistics

 Week 2: Basic concepts of linear algebra

 Week 3: Basics of R

 Week 4: Data visualization

 Week 5: Correlation and regression

 Week 6: Clustering and classification,Correlation and regression

 Week 7: Clustering and classification

 Week 8: Analysis of higher-dimensional data