Course Title: Statistics in Bioinformatics

Course Code CSE- 877

Course Objectives: This is a 3 credit hours course designed for students with a background in biological sciences who have covered the pre-requisite courses of Computational Drug Design and Advanced Computational Biology. The course aims to provide students with the background knowledge of probability distributions and statistical knowledge applicable in analysis of molecular biology data. Furthermore, it will use the R statistical package for the implementation of these concepts. R is a widely used for data analysis and visualization in the scientific community and this will provide students with a useful set of skills to apply in their own research work.

Course Outcomes:

After successfully completing this course, students will gain

- A breadth and depth of multivariate visualization and computing skills necessary to apply high level analytical thinking to complex real-world problems of applied nature
- An exposure to the essentials of computing tools/languages and statistical software use and, hence, the ability to adopt technology know-how demanded in the modern job market
- Hands on practice of dealing classification problems, grouping variables, features engineering, machine learning strategies in both qualitative and quantitative scenarios concerning data driven decision support systems in bioinformatics and associated fields

Detailed Contents:

- Distributions: putting particular focus on Binomial, Negative Binomial, Multinomial,
- Poisson, and Gaussian distributions and their applications in moleuclar biology.
- Data display and descriptive statistics
- Univariate and multivariate statistics
- Hypothesis testing
- Correlation and Regression analysis
- Particularly looking at Regression, Anova, PCA and HCA.
- Linear models.
- Micro Array Analysis
- Analyzing sequences
- Markov models

Recommended / Reference Books:

- Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids by: Richard Durbin, Sean R. Eddy, Anders Krogh, Graeme Mitchison (13 May 1998) Key: citeulike:163532
- R programming for bioinformatics, by Robert Gentleman, Boca Raton, Chapman & Hall/CRC, 2009, ISBN 1-42006-367-7
- Applied Statistics for Bioinformatics using R, by Wim P. Krijnen
- Computer Simulation and Data Analysis in Molecular Biology and Biophysics: An Introduction Using R, by Victor A. Bloomfield, 2009 (Springer)