321 Numerical methods in computational molecular biology
Lecture: TR Phillips 203 1:25-2:15PM,   Section: Hollister B14 R 2:30-3:20
Web site: http://www.cs.cornell.edu/Courses/cs321/2005fa/

Ron Elber, Upson 5148, ron@cs.cornell.edu, Phone 255-7416
Walk-in on Thursdays 10:00, or by appointment
Yonathan Ben Simhon, 304 Upson Hall, bensimho@cs.cornell.edu, Phone 255-5431
Office hours: Tuesday 2:30-3:30 or by appointment

Statistics and Numerical methods motivated by problems from computational molecular biology.
Grades: 5% quiz; 15% homework; 50% projects; 20% final

Textbooks:
M.H. DeGroot and M.J. Schervish, "Probability and Statistics", Addison Wesley
Rudra Pratap, "Getting started with MATLAB 5", Oxford University Press
Charles F. Van Loan, "Introduction to Scientific Computing", Prentice-Hall

(*) Probability: Continuous variables
    1. Random variables
    2. Probability density function
    3. Uniform, normal and exponential distribution
    4. Expectations and variance
    5. Law of large numbers
    6. Central limit theorem
    7. More than one random variables
    8. Rejection and transformation methods for sampling distributions
    9. Transformation methods for sampling distributions (section)
(*) Statistics:
    1. Estimators: Mean, standard deviation
    2. Maximum likelihood
    3. Confidence interval
    4. chi-2 statistics
    5. Regression
    6. Goodness of fit
(*) Matlab preliminaries:
    1.   What is Matlab?
    2.   Sample of simple programs
    3.   Plotting in one and more dimensions
    4.   Writing functions
(*) Analysis of protein shapes
    1.   Plotting protein shapes (vectors, matrices, PDB files)
    2.   Smoothing and interpolation (linear interpolations, linear problems, cubic splines)
    3.   Rotations and overlapping structures (least square fit, Lagrange multipliers, Matrix
         diagonalization, singular value decomposition)
    4.   Clustering protein shapes (building RMS matrices, eigenvalue analysis, identification of families)
(*) Sequence analysis
    1.   Manipulation of characters in Matlab
    2.   Alignment of sequences with gaps
    3.   Dynamic programming
 (*) Optimization
    1.   Steepest Descent
    2.   Conjugate gradient
    3.   Discrete optimization by Monte Carlo: Design of an optimal sequence

    Each student in this course is expected to abide by the Cornell University Code of Academic Integrity.
    Any work submitted by a student in this course for academic credit will be student's own work.