BioE241 labs

374C Stanley Hall, UC Berkeley

Of all natural systems, living matter preserves inscribed in its organization the largest amount of its own past history .... no other system is better aufgehoben: *constantly abolished and simultaneously preserved.* [Pauling and Zuckerkandl, 1963]

Contents

Introduction

This handout describes a series of labs for BioE241, accompanying the theoretical lectures in that class, which describe various statistical models for biological data.

The class has a (non-exclusive) emphasis on models that describe the evolution of DNA, RNA and amino acid sequences on phylogenetic trees.

The BioE241 class page is here: `http://biowiki.org/BioE241`

Most of the labs use the DART software: `http://biowiki.org/DART`

Other software tools are also used. The labs also ask you to develop some pseudocode and actual implementations (generally in a programming language of your choice) and to do a small amount of elementary math.

1 Simulate a discrete-state continuous-time Markov chain on a tree

[Gillespie, 1977]

2 Use profile HMM training and search tools to build a family of homologous protein domain sequences

3 Reconstruct the phylogenetic history of sequences

4 Use probabilistic models to align protein sequences and reconstruct ancestors, on a given phylogeny

5 Simultaneously reconstruct the phylogeny and the alignment

6 Estimate the indel and substitution rates

7 Use probabilistic models to predict the structure of a single RNA sequence

8 Given two related RNA sequences, simultaneously align them and predict their common secondary structure

9 Annotate conserved secondary structure in an RNA multiple alignment and reconstruct ancient RNA sequences

10 Develop and fit models that allow for lineage-specific evolutionary effects

11 Detect recombination breakpoints in multiple sequence alignments

12 Use PRISM to prototype probabilistic models using statistical logic programming

13 Simulate and analyze spatiotemporal models of evolution, epidemiology, ecology, and population dynamics

14 Analyze and fit data to continuous-valued diffusion processes

References

[Gillespie, 1977] Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81:2340–2361.

[Pauling and Zuckerkandl, 1963] Pauling, L. and Zuckerkandl, E. (1963). Chemical paleogenetics, molecular "restoration studies" of extinct forms of life. *Acta Chemica Scandinavica*, 17:S9–S16.