Computational Biology: Combinatorial and Machine Learning Algorithms

2024 • ITSTP

Basic information

Course Title	Machine Learning in Genomics
Instructor	Prof. Sorin Istrail, Department of Computer Science, Brown University.
Prerequisites	Programming: Python
Required Text & Tools	Introduction to Computational Biology: Maps, Sequences and Genomes by Michael Waterman. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids by Richard Durbin, Sean Eddy, Anders Krogh, and Graeme Mitchison.
Grading Criteria	Exam: 50% Homework: 40% Class Participation: 10%
Course Key Words	Sequence Alignment, Finite State Automaton, Knuth-Morris-Pratt, Phylogenetic Tree, Hidden Markov Models (HMMs), Graphical Models, Genome Assembly, Integrative Genomics

Schedule

No.	Topics
Lecture 1	Sequence Alignment: Needleman-Wunsch Algorithm (global alignment); Smith-Waterman Algorithm (local alignment); Multiple Sequence Alignment
Lecture 2	Combinatorial Pattern Matching: Finite Automata; Suffix Tree Algorithms;
Lecture 3	Combinatorial Pattern Matching: Knuth-Morris-Pratt Algorithm; Unsupervised Learning and Clustering: Finding Regulatory DNA Motifs
Lecture 4	Phylogenetic Trees: UPGMA Algorithm (evolutionary distance matrices with uniform clock); Neighbour-Joining Algorithm (general evolutionary distance matrices); Classification Trees and Decision Trees
Lecture 5	Hidden Markov Models (HMMs): Forward Algorithm for PB 1. "Computing the probability" (model scoring); Viterbi Algorithm for PB 2. "Best Explanation" (Viterbi maximum likelihood)

Lecture 6	Hidden Markov Models (HMMs): Graphical Models: Hidden Markov Models and Applications to Bioinformatics
Lecture 7	Genome Assembly: De Brujin Assembly Algorithm (de Brujin graphs and Eulerian paths)
Lecture 8	Integrative Genomics: Multi-Omics Data Fusion
Lecture 9	Protein Folding Algorithms (Introduction)
Lecture 10	Bioinformatics Analysis of the Coronavirus Genome: All the algorithms presented in previous lectures are of basic importance in the analysis of the coronavirus genome.