

Algorithmische Bioinformatik II (IN5001)

Title	Algorithmic Bioinformatics II	
Typ	Vorlesung mit Übungen	
Credits	9	
Lehrform/SWS	4V + 2Ü	
Sprache	Deutsch oder Englisch (wird vom Dozenten zum Vorlesungsbeginn bekannt gegeben)	
Modulniveau	Bachelor	
Arbeitsaufwand	Präsenzstunden	90 Stunden
	Eigenstudium	180 Stunden
	Gesamtaufwand	270 Stunden
Angestrebte Lernergebnisse	Die Teilnehmer sind in der Lage, weiterführende Methoden der Bioinformatik anzuwenden und zu analysieren. Insbesondere sind die Studierenden in der Lage, bioinformatische Probleme bzgl. der NP-Vollständigkeit und der Komplexitätsklassen für Optimierungsprobleme einzuordnen, einfache approximative und heuristische Algorithmen zu entwerfen und zu analysieren, Algorithmen zum mehrfachen Sequenzen-Alignment zu entwickeln und für ein gegebenes bioinformatisches Problem probabilistische Modelle zu entwerfen.	
Intended Learning Outcomes	Students are able to apply and analyze advanced bioinformatics methods. In particular students are able to classify Bioinformatics problems in NPC as well as in complexity classes for optimization problems, to design and analyze approximation and heuristic algorithms, to develop algorithms for multiple sequence alignment, to construct a probabilistic model for a given Bioinformatics problem.	
Inhalt	Themen des Moduls Algorithmische Bioinformatik II sind: <ul style="list-style-type: none"> • NP-Vollständigkeit: Reduzierbarkeit, NP-vollständige Probleme in der Bioinformatik • Heuristische und approximative Ansätze (PTAS, APX) • Randomisierte Algorithmen • Mehrfaches Sequenzen-Alignment (n-dimensionale DP; heuristische Verfahren; approximative Methoden; Baum-, Consensus-, Profil- und Center-Star-Verfahren; DiAlign; ClustalW) • Probabilistische Modellierung (Orthodoxe und Bayessche Modellierung; Bayessche Inferenz; 	

	<p>Optimierung mit constraints; Algorithmen: EM, MC, MCMC, Gibbs Sampling)</p> <ul style="list-style-type: none"> • Markov-Ketten und Hidden-Markov-Modelle (HMMs) (Modelle; Hauptprobleme; Algorithmen: Forward/Backward, Viterbi, Baum-Welch (EM)) • Optional: Clustering Modelle und Algorithmen (CLICK: Graph Partitionierung, probabilistische Graph Modelle) • Optional: Expressiondatenanalyse (Image analysis, Präprozessierung, Normalisierung, Centralization, (Un-/)Supervised Methoden)
<p>Contents</p>	<p>Topics of Algorithmic Bioinformatics II are:</p> <ul style="list-style-type: none"> • NP-completeness: reductions, NP-complete problems in bioinformatics • heuristic and approximation algorithms (PTAS, APX) • randomized algorithms • multiple alignment (n-dimensional DP; heuristic approaches; approximations; tree, consensus, profile and center star methods; DiAlign; ClustalW) • probabilistic modeling (orthodox and Bayesian modeling; Bayesian inference; optimization with constraints; EM, MC, MCMC, Gibbs sampling) • Markov chains and hidden Markov models (HMMs) (models; main problems; algorithms: forward/backward, Viterbi, Baum-Welch (EM)) • Optional: clustering models and algorithms (CLICK: graph partitioning, probabilistic graph model) • Optional: Expression data analysis (image analysis, preprocessing, normalization, centralization, (Un-/)Supervised methods)
<p>Prüfung</p>	<p>Prüfungsleistung (benotet): Klausur (120min)</p> <p>Wiederholungsklausur zu Ende des Semesters. Details werden zu Beginn des Moduls bekannt gegeben.</p> <p>In der Klausur weisen die Studierenden nach, dass sie bioinformatische Probleme bzgl. ihrer Zeitkomplexität klassifizieren, den Ablauf der im Modul behandelten Algorithmen erläutern und diese Algorithmen anwenden sowie einfache approximative und heuristische Algorithmen entwerfen und bzgl. ihrer</p>

	<p>Laufzeit analysieren können. Des Weiteren können sie für einfache bioinformatische Problemstellungen probabilistische Modelle erstellen. Konkret werden in der Klausur 4-7 Aufgaben bearbeitet, die eine eigenständige Anwendung bioinformatischer Methoden und Modelle zur Lösung einer anspruchsvollen Problemstellung erfordern, wie die konkrete Komplexitätstheoretische Klassifizierung eines Problems (z.B., in NPO, APX, PTAS samt zugehöriger Reduktion), die Beschreibung des Verhaltens eines im Modul behandelten Algorithmus (z.B. MSA, Center-Star, MCMC, EM, Viterbi, BW), das Anpassen und Entwickeln eines einfachen approximativen oder heuristischen Algorithmus (auch für das mehrfache Sequenzen-Alignment) und die probabilistische Modellierung eines gegebenen Problems (z.B. Inferenz von Modellparametern, Markov-Modell, HMM).</p>
Examination	<p>Examination requirements (graded): written exam (120 min)</p> <p>A makeup exam will be offered at the end of the semester. Details will be announced at the beginning of the module.</p> <p>Within the written exam, students demonstrate that they are able to classify Bioinformatics Problems w.r.t. complexity classes for optimization problems, to explain the behavior of algorithms presented in the module for specific data and to apply these algorithms, to develop and to analyze simple approximation and heuristic algorithms. Furthermore, they are able to create probabilistic models for simple Bioinformatics problems. The written exam contains 4 to 7 assignments, which require independent application of methods and models (presented in the module) to solve demanding problems such as the complexity-theoretic classification of problems (e.g., NPO, APX, PTAS including the corresponding reductions), the description of the behavior of algorithms presented in the module (e.g., MSA, Center-Star, MCMC, EM, Viterbi, BW), the adaptation and design of a simple approximation or heuristic algorithm (including algorithms for multiple sequence alignment) as well as the probabilistic modeling of a given problem (e.g., inference of model parameters, Markov models, HMM).</p>
Literatur/Literature	S. Aluru (Ed.): Handbook of Computational Molecular

	<p>Biology, Chapman and Hall/CRC, 2006.</p> <p>H.-J. Böckenhauer, D. Bongartz: Algorithmische Grundlagen der Bioinformatik: Modelle, Methoden und Komplexität, Teubner, 2003.</p> <p>P. Clote, R. Backofen: Computational Molecular Biology - An Introduction, Wiley, 2000.</p> <p>R.C. Deonier, S. Tavaré, M.S. Waterman: Computational Genome Analysis, Springer, 2005.</p> <p>R. Durbin, S. Eddy, A. Krogh, G. Mitchinson: Biological Sequence Analysis - Probabilistic Models of Proteins and Nucleic Acids, Cambridge University Press, 1998.</p> <p>D. Gusfield: Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology, Cambridge University Press, 1997.</p> <p>V. Heun: Algorithmische Bioinformatik, Skriptum, 2001-2016.</p> <p>N.C. Jones, P.A. Pevzner: An Introduction to Bioinformatics Algorithms, MIT Press, 2004.</p> <p>D.W. Mount: Bioinformatics - Sequence and Genome Analysis, Cold Spring Harbor Laboratory Press, 2001.</p> <p>P. Pevzner: Computational Molecular Biology - An Algorithmic Approach, MIT Press, 2000.</p> <p>J.C. Setubal, J. Meidanis: Introduction to Computational Molecular Biology, PWS Publishing Company, 1997.</p> <p>M.S. Waterman: Introduction to Computational Biology: Maps, Sequences, and Genomes, Chapman and Hall, 1995.</p>
Medienformen	Folienpräsentation, Tafelanschrieb
Media	slide show, blackboard
Lehr- und Lernmethode	<p>Vorlesung, Tutorübung, Aufgaben zum Selbststudium.</p> <p>Das Modul besteht aus einer Vorlesung sowie Übungen in kleinen Gruppen. In den Hausaufgaben, die freiwillig abzugeben sind, wird das Verständnis der Methoden, Modelle und Algorithmen, die in der Vorlesung vorgestellt werden, anhand konkreter Daten und Beispiele vertieft. Die Studierenden entwerfen mithilfe der in der Vorlesung vorgestellten Methoden und Modelle probabilistische Modelle sowie neue Algorithmen bzw. adaptieren bekannte Algorithmen für ähnliche Probleme und analysieren diese. In den Hausaufgaben werden selbständig anspruchsvolle Übungsaufgaben bearbeitet, die ähnlich wie die Klausuraufgaben sind (siehe oben) und daher zur Vorbereitung darauf dienen. In den Übungen werden mögliche Lösungsansätze der Aufgaben zum Selbststudium diskutiert.</p>
Teaching and Learning	Lecture, tutorial, assignments for individual study.

Methods	<p>The module consists of a lecture and in addition exercises in small groups. Within the assignments (submission is optional), methods, models, and algorithms (presented in the lecture) will be applied to real data and examples to deepen the understanding. The students develop probabilistic models as well as new algorithms based on the presented methods and techniques or adapt known algorithms to related problems and analyze these algorithms. The assignments consist of demanding problems similar to the assignments in the written exam (for details see above) and serve as a preparation for the written exam. Within the tutorials possible approaches for solutions of the assignments will be discussed.</p>
Turnus	Wintersemester
Modulverantwortlicher	Prof. Dr. Ralf Zimmer
Dozenten	Prof. Dr. Caroline Friedel Prof. Dr. Volker Heun Prof. Dr. Ralf Zimmer