COURSE DESCRIPTION CARD - SYLLABUS

Course name
Statistical Methods in Structural Bioinformatics [S2Bioinf1>MSTAT]

Course

Field of study Bioinformatics	Year/Semester 2/3
Area of study (specialization) –	Profile of study general academic
Level of study second-cycle	Course offered in polish
Form of study full-time	Requirements elective

Number of hours

Lecture 15	Laboratory classes 15	Other (e.g. online) 0
Tutorials 0	Projects/seminars 0	

Number of credit points

2,00

Coordinators	Lecturers
dr hab. inż. Tomasz Żok tomasz.zok@put.poznan.pl	

Prerequisites

A student entering this course should have a basic knowledge of structural biology, programming, and statistical analysis. He or she should also have the ability to obtain information from external sources, and should have patience, diligence, and curiosity.

Course objective

To teach students the theory and show practical solutions in structural bioinformatics involving the use of statistical methods. To develop skills of modeling problems and solving them using the learned methods.

Course-related learning outcomes

Knowledge:
1. Knows the basic formats for storing biological structures and methods for their processing.
2. Knows the theoretical background of selected statistical models used in structural bioinformatics.
3. Knows existing software for statistical analysis of protein and nucleic acid structures.
4. Knows programming libraries containing functions for statistical analysis.

Skills:

1. Is able to analyze selected sets of structured data using statistical methods.
2. Is able to perform selected statistical tests and correctly interpret their results.
3. Is able to develop software to solve selected structural bioinformatics problems.
4. Is able to design and perform selected computational experiments and prepare a report of the results.

Social competences:
1. Is able to prioritize and plan work accordingly.
2. Understands the need for constant learning of new things.
3. Understands the dangers of incorrect statistical inference and the responsibility involved.

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

Learning outcomes presented above are verified as follows:
Knowledge acquired in the lecture is verified by a colloquium during the last lecture in the semester. The test consists of 10-15 questions (closed or open), with different numbers of points. Pass mark: 50% of the points. Examination topics will be presented during the lectures.
The skills acquired in laboratory classes will be verified on the basis of the evaluation of projects carried out by the students during the semester. The project will be evaluated depending on the complexity of the problem (students will be able to work on basic or complex problems), the quality of the prepared code and report, as well as the presentation during which the student will present the results obtained.

Programme content

The course consists of seven lectures. The first will cover the theoretical foundations of the methods presented later: full and simplified Bayesian inference models, Boltzmann distribution, free energy, information theory and entropy. This knowledge will allow to present practically used methods such as Bayesian networks and hidden Markov models. Further lectures will show real-world applications of these models in structural bioinformatics for (1) determining the energy of a structure and thus assessing its quality, (2) analyzing distributions of angular values describing the spatial shape of proteins and RNAs, and (3) probabilistic sampling of structural fragments.
The laboratory exercises are closely related to the topics of the lectures. Theoretical knowledge acquired in this way will be applied to practical examples. Students will get acquainted with the provided materials and will perform the tasks recommended in each class. For their implementation they will use real biological data deposited in publicly available databases. Students will become familiar with existing software for statistical structural analysis including its advantages, disadvantages and limitations. At the same time, students will design and implement their own solutions, whose usefulness will be verified by conducting computational experiments.

Teaching methods

1. Lecture: multimedia presentation
2. Laboratory classes: multimedia presentation, practical exercises, performing computational experiments, preparing reports

Bibliography

Basic
1. T. Hamelryck, K. Mardia, J. Ferkinghoff-Borg, „Bayesian Methods in Structural Bioinformatics”
2. R. Durbin „Biological Sequence Analysis (Probabilistic Models of Protein and Nucleic Acids)
Additional
1. N. Fisher, „Statistical Analysis of Circular Data”
2. K. Mardia, „Directional Statistics”
3. S. Brooks, A. Gelman, G. L. Jones, X.-L. Meng, „Handbook of Markov Chain Monte Carlo”

Breakdown of average student's workload

	Hours	ECTS
Total workload	50	2,00
Classes requiring direct contact with the teacher	30	1,00
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	20	1,00