



## KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Metody statystyczne w bioinformatyce strukturalnej [S1Bioinf1>MSTAT]

### Przedmiot

Kierunek studiów  
Bioinformatyka

Rok/Semestr  
4/7

Studia w zakresie (specjalność)  
–

Profil studiów  
ogólnoakademicki

Poziom studiów  
pierwszego stopnia

Język oferowanego przedmiotu  
polski

Forma studiów  
stacjonarne

Wymagalność  
obieralny

### Liczba godzin

Wykład  
15

Laboratorium  
15

Inne (np. online)  
0

Ćwiczenia  
0

Projekty/seminaria  
0

### Liczba punktów ECTS

2,00

### Koordynatorzy

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

### Wykładowcy

### Wymagania wstępne

Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę o biologii strukturalnej, programowaniu oraz analizie statystycznej. Powinien również posiadać umiejętność pozyskiwania informacji ze źródeł, w szczególności w języku angielskim, oraz wykazywać się cierpliwością, starannością i ciekawością poznawczą.

### Cel przedmiotu

Przekazanie studentom teorii oraz przedstawienie praktycznych rozwiązań w bioinformatyce strukturalnej polegających na wykorzystaniu metod statystycznych. Rozwijanie umiejętności modelowania problemów i ich rozwiązywania z użyciem poznanych metod.

### Przedmiotowe efekty uczenia się

Wiedza:

1. Zna podstawowe formaty zapisu struktur biologicznych oraz metody ich przetwarzania.
2. Zna teoretyczne podstawy wybranych modeli statystycznych wykorzystywanych w bioinformatyce strukturalnej.
3. Zna istniejące oprogramowanie do statystycznej analizy struktur białek i kwasów nukleinowych.

4. Zna biblioteki programistyczne z funkcjami do analizy statystycznej.

Umiejętności:

1. Umie przeanalizować wybrane zbiory danych strukturalnych z wykorzystaniem metod statystycznych.
2. Umie wykonać wybrane testy statystyczne oraz poprawnie zinterpretować ich wyniki.
3. Umie stworzyć oprogramowanie komputerowe do rozwiązywania wybranych problemów bioinformatyki strukturalnej.
4. Umie zaprojektować i wykonać wybrane eksperymenty obliczeniowe oraz przygotować raport z wynikami.

Kompetencje społeczne:

1. Potrafi ustalać priorytety i odpowiednio zaplanować pracę.
2. Rozumie potrzebę uczenia się przez całe życie.
3. Rozumie zagrożenia wynikające z nieprawidłowego wnioskowania statystycznego i wiążącą się z tym odpowiedzialność.

### Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Wiedza nabyta w ramach wykładu weryfikowana jest przez kolokwium podczas ostatniego wykładu w semestrze. Kolokwium składa z 10-15 pytań (testowych i otwartych), różnie punktowanych. Próg zaliczeniowy: 50% punktów. Zagadnienia zaliczeniowe przedstawione zostaną w trakcie wykładów. Umiejętności nabyte w ramach zajęć laboratoryjnych weryfikowane będą na podstawie oceny projektów wykonywanych przez studentów w trakcie semestru. Na ocenę projektu wpływa złożoność podjętego problemu (studenci będą mogli pracować nad podstawowymi lub rozszerzonymi zagadnieniami), jakość przygotowanego kodu oraz sprawozdania, a także prezentacja w trakcie, której student będzie referował uzyskane wyniki.

### Treści programowe

- Pełne i uproszczone modele wnioskowania bayesowskiego
- Rozkład Boltzmanna
- Energia swobodna
- Teoria informacji i entropia
- Sieci bayesowskie i ukryte modele Markowa
- Określanie energii struktury i ocena jej jakości
- Analiza rozkładów wartości kątowych opisujących przestrzenny kształt białek i RNA
- Probabilistyczne próbkowanie fragmentów strukturalnych

### Tematyka zajęć

Kurs składa się z siedmiu wykładów. Pierwsze z nich dotyczyć będą teoretycznych podstaw prezentowanych później metod: pełne i uproszczone modele wnioskowania bayesowskiego, rozkład Boltzmanna, energia swobodna, teoria informacji i entropia. Wiedza ta pozwoli na przedstawienie praktycznie wykorzystywanych metod takich jak sieci bayesowskie oraz ukryte modele Markowa. Dalsze wykłady pokazywać będą rzeczywiste zastosowania tych modeli w bioinformatyce strukturalnej do (1) wyznaczania energii struktury i tym samym oceny jej jakości, (2) analizy rozkładów wartości kątowych opisujących przestrzenny kształt białek i RNA, oraz (3) probabilistycznego próbkowania fragmentów strukturalnych.

Ćwiczenia laboratoryjne związane są ściśle z tematyką wykładów. Teoretyczna wiedza nabyta w ten sposób będzie wykorzystana na praktycznych przykładach. Studenci zapoznają się z dostarczonymi materiałami oraz wykonują polecane na każdym zajęciach zadania. Do ich realizacji wykorzystują rzeczywiste dane biologiczne zdeponowane w publicznie dostępnych bazach danych. Studenci poznają na ćwiczeniach istniejące oprogramowanie do statystycznej analizy strukturalnej z uwzględnieniem jego zalet, wad i ograniczeń. Jednocześnie studenci w ramach zadań zaprojektują i zaimplementują własne rozwiązania, których użyteczność zweryfikują, przeprowadzając eksperymenty obliczeniowe.

### Metody dydaktyczne

1. Wykład: prezentacja multimedialna
2. Ćwiczenia laboratoryjne: prezentacja multimedialna, ćwiczenia praktyczne, wykonywanie

eksperymentów obliczeniowych, przygotowywanie sprawozdań

## Literatura

Podstawowa

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)

Uzupełniająca

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”

## Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	50	2,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	30	1,00
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwium/egzaminu, wykonanie projektu)	20	1,00