AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY

Module name: Hands-on introduction to biomedical data analysis in Python

Academic year: 2019/2020 Code: ZSDA-3-0026-s ECTS credits: 5

Faculty of: Szkoła Doktorska AGH

Field of study: Szkoła Doktorska AGH Specialty: —

Study level: Third-cycle studies Form and type of study: Full-time studies

Lecture language: English Profile of education: Academic (A) Semester: 0

Course homepage: —

Responsible teacher: dr inż. Surówka Artur (Artur.Surowka@fis.agh.edu.pl)

Module summary

The course aims at introducing a PhD student to simple and flexible means for critically analyzing data and to assure their proper presentation. The course is devoted to students from the wide range of life science: physics, chemistry, biophysics, biology, biotechnology...who feel the need for gaining basic practical skills in programming, data analysis and presentation.

In the case of any enquires, do not hesitate to write an email to: asurowka@agh.edu.pl

Description of learning outcomes for module

MLO code	Student after module completion has the knowledge/ knows how to/is able to	Connections with FLO	Method of learning outcomes verification (form of completion)
Social competence: is able to			
M_K001	A student can address questions, seeking support and interact with the lecturer and group members during lectures and tutorials in English.	SDA3A_U02, SDA3A_U05, SDA3A_U01	Activity during classes
Skills: he can			
M_U001	A student can prepare the presentation and showcase their results in English to the interdisciplinary audience.	SDA3A_U03, SDA3A_U02, SDA3A_U05, SDA3A_U01, SDA3A_U04	Presentation
M_U002	A student can choose proper methods for data pre-processing	SDA3A_U03, SDA3A_U02, SDA3A_U01	Presentation, Oral answer, Activity during classes

M_U003	A student can recognize major data types and extensions and find proper methods for acessing them and saving	SDA3A_U03, SDA3A_U01	Oral answer, Activity during classes
M_U004	A student can critically judge the data quality and design the most appropriate data analysis pipeline	SDA3A_U03, SDA3A_U02, SDA3A_U05, SDA3A_U01, SDA3A_U04	Oral answer, Activity during classes
M_U005	A student can find the best means for data presentation, and can identify the most common pitfalls	SDA3A_U07, SDA3A_U03, SDA3A_U02, SDA3A_U05, SDA3A_U01, SDA3A_U04	Presentation, Oral answer, Activity during classes
M_U006	A student can use basic Python syntax to the extent necessary to perform basic data programming tasks	SDA3A_U07, SDA3A_U06, SDA3A_U03, SDA3A_U02, SDA3A_U05, SDA3A_U01, SDA3A_U04	Oral answer, Activity during classes
Knowledge: he knows and understands			
M_W001	A student knows basic terminology used in computer programming.	SDA3A_W03, SDA3A_U02, SDA3A_U05, SDA3A_U01	Activity during classes, Oral answer
M_W002	A student knows basic methods and tools for accessing data.	SDA3A_W03, SDA3A_U03	Oral answer, Activity during classes
M_W003	A student knows the theoretical background of the major methods for multivariate data analysis.	SDA3A_W03	Oral answer, Activity during classes
M_W004	A student knows basic tools for spectral analysis.	SDA3A_W03	Oral answer, Activity during classes
M_W005	A student knows major means and rules for proper data visualization.	SDA3A_W03, SDA3A_W04, SDA3A_W06	Presentation, Oral answer, Activity during classes
M_W006	A student knows basics of syntax in Python	SDA3A_W03	Oral answer, Activity during classes

Number of hours for each form of classes

Suma	Form of classes										
	Lectures	Auditorium classes	Laboratory classes	Project classes	Conversation seminar	Seminar classes	Practical classes	Fieldwork classes	Workshops	Prace kontrolne i przejściowe	Lektorat
75	30	0	30	0	0	15	0	0	0	0	0

FLO matrix in relation to forms of classes

MLO code	Student after module completion has the knowledge/ knows how to/is able to	Form of classes

		Lectures	Auditorium classes	Laboratory classes	Project classes	Conversation seminar	Seminar classes	Practical classes	Fieldwork classes	Workshops	Prace kontrolne i przejściowe	Lektorat
Social competence: is able to												
M_K001	A student can address questions, seeking support and interact with the lecturer and group members during lectures and tutorials in English.	+	-	+	-	-	+	-	-	-	-	-
Skills: he can												
M_U001	A student can prepare the presentation and showcase their results in English to the interdisciplinary audience.	-	-	-	-	-	+	-	-	-	-	-
M_U002	A student can choose proper methods for data pre-processing	-	-	+	-	-	+	-	-	-	-	-
M_U003	A student can recognize major data types and extensions and find proper methods for acessing them and saving	-	-	+	-	-	+	-	-	-	-	-
M_U004	A student can critically judge the data quality and design the most appropriate data analysis pipeline	+	-	+	-	-	+	-	-	-	-	-
M_U005	A student can find the best means for data presentation, and can identify the most common pitfalls	+	-	+	-	-	+	-	-	-	-	-
M_U006	A student can use basic Python syntax to the extent necessary to perform basic data programming tasks	+	-	+	-	-	+	-	-	-	-	-
Knowledge: he knows and understands												
M_W001	A student knows basic terminology used in computer programming.	+	-	+	-	-	+	-	-	-	-	-
M_W002	A student knows basic methods and tools for accessing data.	+	-	+	-	-	+	-	-	-	-	-
M_W003	A student knows the theoretical background of the major methods for multivariate data analysis.	+	-	+	-	-	+	-	-	-	-	-
M_W004	A student knows basic tools for spectral analysis.	+	-	+	-	-	+	-	-	-	-	-
M_W005	A student knows major means and rules for proper data visualization.	+	-	+	-	-	+	-	-	-	-	-
M_W006	A student knows basics of syntax in Python	+	-	+	-	-	+	-	-	-	-	-

Student workload (ECTS credits balance)

Student activity form	Student workload
Udział w zajęciach dydaktycznych/praktyka	75 h
Preparation for classes	15 h
Preparation for classes	15 h
przygotowanie projektu, prezentacji, pracy pisemnej, sprawozdania	10 h
przygotowanie projektu, prezentacji, pracy pisemnej, sprawozdania	10 h
Realization of independently performed tasks	15 h
Realization of independently performed tasks	15 h
Examination or Final test	2 h
Contact hours	5 h
Contact hours	5 h
Summary student workload	167 h
Module ECTS credits	5 ECTS

Additional information

Module content

Lectures

Lectures’ programme

1) Practical introduction to programming in Python 3.6 (6h).
2) Handling scientific data: basic data types, readout and saving (2h).
3) Data pre-processing techniques (4h).
4) Data analysis:
- Basics of spectral analysis (4h)
- Basics of digital image analysis and quantitative morphometry (2h).
- Statistical testing (4h).
- Multivariate modeling methods (4h).
5) Visualisation, presentation and reporting on scientific data (6h)

1. 2. 3. 4.

Program wykładów:

1) Praktyczne wprowadzenie do programowania w języku Python 3.6 (6h).
2) Przetwarzanie danych naukowych: główne typy danych, odczyt i zapis (2h).
3) Techniki wstępnego przetwarzania danych (4h)
4) Analiza danych:
- podstawy analizy spektroskopowej (4h),
- podstawy analizy obrazów cyfrowych i morfometrii ilosciowej (2h),
- wnioskowanie statystyczne (4h),
- metody modelowania wielowymiarowego (4h).

5) Wizualizacja, prezentacja i raportowanie danych naukowych (4h).

Laboratory classes

Hands-on tutorials' programme:

1) Loops, conditional expressions and error handling (2h).
2) Major data structures: lists, tuples, arrays and data frames (2h).
3) Functions and basics of classes (2h).
4) Data access: reading/saving (2h).
5+6) Data transformation and exploration methods (2h + 2h).
7) Basics of spectral analysis: filtration, normalization tools, integration (2h) and fitting (2h).
8) Basics of digital image analysis and quantitative morphometry: image filtration, bizarization and major parameters describing shape (2h).
9) Parametric testing methods (2h).
10) Non-parametric testing methods (2h).
11) Multivariate modeling methods: PCA, LDA, PLS, HCA (2h).
12) Data visualization, presentation (4h) and reporting (2h).

1. 2. 3. 4.

Program praktycznych ćwiczeń laboratoryjnych

1) Pętle, instrukcje warunkowe, sterowanie wyjątkami (2h)
2) Główne struktury danych: listy, krotki, tablice i ramki (2h)
3) Funkcje i podstawowe informacje o klasach (2h)
4) Dostęp do danych: otwieranie/zapisywanie danych (2h)
5 + 6) Techniki transformacji i eksploracji danych (2h + 2h)
7) Podstawy analizy spektroskopowej: filtracja, normalizacja, całkowanie (2h) i dopasowanie (2h).
8) Podstawy analizy obrazów cyfrowych: filtracja, binaryzacja i ekstrakcja parametrów kształtu (2h).
9) Parametryczne testy statystyczne (2h)
10) Nieparametryczne testy statystyczne (2h)
11) Wielowymiarowe metody modelowania: PCA, LDA, PLS, HCA (2h).
12) Wizualizacja, prezentacja (4h) i opis danych naukowych (2h)

Seminar classes

The seminars will give to a student an opportunity to take advantage of the already acquired programming skills and discuss their own scientific results to the interdisciplinary audience by using the methods presented (short presentation, max 10 min + 5 mins for discussion – down to the number of attendees) during the whole course. This activity is planned to be blocked into two longer meetings.

1. 2. 3. 4.

Seminaria dadzą studentowi okazję do wykorzystania i zweryfikowania nabytych umiejętności programistyczno-analitycznych, jak również stworzą możliwość do zaprezentowania swoich własnych wyników naukowych w interdyscyplinarnym gronie (krótka prezentacja max 10 minut plus 5 minut na dyskusję – czasy zależny od liczby uczestników kursu).

Teaching methods and techniques:
Lectures: – prezentacja multimedialna
Laboratory classes: – praca z komputerem
Seminar classes: – przygotowanie prezentacji wyników badan (własnych), w oparciu o nabyte umiejętności

Warunki i sposób zaliczenia poszczególnych form zajęć, w tym zasady zaliczeń poprawkowych, a także warunki dopuszczenia do egzaminu:
In liaison with the University regulations.

Lectures: attending min 70% of lectures.
Laboratories: active participation and compketing most of the planned computer exercises.
Seminars: presenting the data analyzed by the methods acquired during the course.

1. 2. 3.

Wg regulaminu AGH.

Wykłady: min 70% obecności.
Laboratoria: aktywny udział i prawidłowe wykonanie większości ćwiczeń.
Seminaria: zaprezentowanie wyników analizy danych, przeprowadzonej z wykorzystaniem metod poznanych w czasie kursu.

Zasady udziału w poszczególnych zajęciach, ze wskazaniem, czy obecność studenta na zajęciach jest obowiązkowa:
Lectures:
– Attendance is mandatory: Yes
– Participation rules in classes: Nie określono
Laboratory classes:
– Attendance is mandatory: Yes
– Participation rules in classes: Nie określono
Seminar classes:
– Attendance is mandatory: Yes
– Participation rules in classes: Nie określono

Method of calculating the final grade
The grade will only be given when the student is granted the lectures' credit (70% lectures attended)

1) The whole course:

FG = 0.65 L + 0.35 S

L – the final grade achieved for student's activity during the computer laboratories.
S – the final grade for the seminar presentation.

2) The course without the seminars:

FG = L

1. 2. 3. 4. 5.

Ocena końcowa (OK) wystawiona zostanie wyłącznie w przypadku uzyskania zaliczenia wykładów (obecność na 70% wykładów).

1) Cały kurs

OK = 0.65 L + 0.35 S

L – ocena za aktywność studenta w czasie ćwiczeń laboratoryjnych.
S – ocena za prezentację wyników w czasie seminarium.

2) kurs bez seminariów

OK = L

Sposób i tryb wyrównywania zaległości powstałych wskutek nieobecności studenta na zajęciach:

As the tutorials will be strongly related to the content presented during the lectures, the student is expected to attend at least 70% of the lectures to be granted a credit.

A student is allowed maximum two absences the tutorials.

A student is allowed maximum one absence during the seminars.

1. 2. 3. 4.

Jako że ćwiczenia laboratoryjne będą ściśle związane z materiałem prezentowanym na wykładach, student powinien być obecny na 70% wykładów aby uzyskać ich zaliczenie.

Student ma prawo do max dwóch nieobecności na ćwiczeniach laboratoryjnych.

Student ma prawo do max jednej nieobecności na zajęciach seminaryjnych.

Prerequisites and additional requirements

Basic knowledge of mathematics and statistics. Basic programming skills are welcome but not mandatory.

1. 2. 3. 4.

Podstawowa wiedza z zakresu matematyki i statystyki. Podstawowe umiejętności z zakresu programowania są mile widziane ale nie są konieczne,

Recommended literature and teaching resources

1) A. Stanisz. Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Krakow 2006.
2) S. Madhavan. Mastering Python for Data Science. Pact Publishing, Birmingham 2015
3) I. Milovanović, D. Foures, G. Vettigli. Python Data Visualization Cookbook. Pact Publishing, Birmingham 2015

Scientific publications of module course instructors related to the topic of the module

1) A.D. Surowka et al. The combination of artificial neural networks and synchrotron radiation-based infrared micro-spectroscopy for a study on the protein composition of human glial tumors. Analyst. 2015 Apr 7;140(7):2428-38. doi: 10.1039/c4an01867b

2) A.D. Surowka et al. FTIR imaging of the molecular burden around Aβ deposits in an early-stage 3-Tg-APP-PSP1-TAU mouse model of Alzheimer’s disease. Analyst, 2017,142, 156-168. doi:10.1039/C6AN01797E

3) A.D. Surowka et al. Combined in-situ imaging of structural organization and elemental composition of substantia nigra neurons in the elderly. Talanta. 2016 Dec 1;161:368-376. doi: 10.1016/j.talanta.2016.08.023. Epub 2016 Aug 5.

4) A.D. Surowka et al. Novel approaches for correction against the soft matrix effects in the quantitative elemental imaging of human substantia nigra tissue using synchrotron X-ray fluorescence. Spectrochimica Acta Part B: Atomic Spectroscopy. Volume 123, 1 September 2016, Pages 47-58.

5) A.D. Surowka et al. Peripheral Vagus Nerve Stimulation Significantly Affects Lipid Composition and

Protein Secondary Structure Within Dopamine-Related Brain Regions in Rats. NeuroMolecular Medicine June 2015, Volume 17, Issue 2, pp 178–191

6) A.D. Surowka et al. A METHODOLOGICAL APPROACH TO THE CHARACTERIZATION OF BRAIN GLIOMAS, BY MEANS OF SEMI-AUTOMATIC MORPHOMETRIC ANALYSIS. Image Anal Stereol 2014;33:201-218.

7) A.D. Surowka et al. Combined use of infrared and hard X-ray microprobes for spectroscopy-based neuroanatomy. 2018 JINST 13 C05008

Additional information

The course is intended to start in the summer semester 2020.

There may also be a possibility to do just lectures and tutorials for 4 ECTS.

1. 2. 3. 4. 5.

Planowany początek kursu – semestr letni 2020.

Może być możliwość realizacji wyłącznie wykładu i ćwiczeń za 4 ECTS.