UNIVERSITY OF LONDON

BSc Computer Science

Module Specification

Key Information			
Module title	Programming with Data		
Level	5	Credit value	15
Member Institution	Goldsmiths	Notional study hours and duration of course	150
Module lead author/ Subject matter expert	Aaron Gerow		
Module co-author			

Rationale for the module

Increasingly, computer systems in research and industry are designed to leverage large amounts of data. The data explored by such systems are rich and various: they could include anything from the results of clinical trials to information gleaned from analysing millions of tweets to understanding how people talk positively and negatively about politics. In this module, you will learn how to develop systems that operate in, and make use of such data-rich environments. This module builds on other material in the programme such as mathematics and databases.

Aims of the module

This module will show you how to work with data: getting data from a variety of sources, visualising data in compelling, informative ways, processing data to make it useful and shareable, and reasoning with data to test hypotheses and make parameterised predictions. The module will also introduce you to a new language and programming environment that is well-adapted to languages for these applications.

Topics covered in this module:

The topics listed here are an approximation of what will be covered. The topics presented may be slightly revised to ensure currency and relevance. Students will be advised of any changes in advance of their study.

1. Setting up the programming environment
2. Control structures, functions and comprehensions.
3. Data-driven programmming
4. Visualising data
5. Descriptive statistics
6. Getting data
7. Processing data: cleaning, normalizing, and scaling
8. Classification with K-nearest neighbours
9. Bayes' theorem and naïve Bayes classification
10. Clustering

Learning outcomes for the module

Students who successfully complete this module will be able to:

1. Be able to set up and work within a Python environment for writing data-driven software
2. Write simple programs in Python that perform data analysis and visualisation
3. Explain how to retrieve data from various sources and be able to load it into a program
4. Prepare and summarize data
5. Reason about data to make predictions and classifications

Assessment strategy, assessment methods

Summative and Formative Assessments
The module will contain a range of summative and formative assessments. Summative assessments are assessments which contribute directly towards your final grade. Formative assessments do not count directly towards your final grade. Instead, they provide you with opportunities for low stakes practice, and will often provide some sort of feedback about your progress. For example, a practice quiz might provide you with feedback about why a particular answer was wrong.

Assessment Activities
The table below lists the assessment activity types you might encounter taking the module. It also states if that type of assessment can be automatically graded. For example, multiple choice quizzes can be automatically graded, and so can some programming assignments. It also states if that type of assessment will be found in the summative courseworks and the summative exam. More details about the summative assessments are provided below.

Assessment activity type	Can it be automatically graded with feedback in some cases?	CW	Exam
Quiz	X	X	X
Writing task		X	X
Programming task	X	X	X

Pass Mark

In order to pass this module, you must achieve at least 35% in each element of summative assessment and an overall weighted average of 40%, subject to the application of rules for compensation. Please refer to the programme regulations for more information.

Assessment Elements

As this is a module that has a significant amount of theory it is assessed as a theory-based module. This means that the summative assessment is composed of two elements, whose weightings are listed in the table below.

Summative Assessment Component	Percentage of final credit	Deadline
Coursework	50%	Mid session
Examination	50%	End of session

The coursework comprises a variety of practical exercises and quizzes which in total will take up to 25 hours of study time to complete. The examination will be two hours long, and consist of written answer and multiple choice questions.

Learning resources

The module will draw on a number of different, largely web-based, public resources as well as the resources produced as bespoke material for this module. The standard text book(s) for the module will be:

Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, Wes McKinney, O'Reilly Media, ISBN-10: 1491957662, 2017

Online resources include:

Ipython website: https://ipython.org/
Python main portal: https://www.python.org/