Foundations of Data Science Course Syllabus

Course	Foundations of Data Science	Faculty	Philosophy
Course Code	NCHDS442	Course Leader	Dr Alexandros Koliousis
Credit Points	15	Teaching Period	Hilary
FHEQ Level	Level 4	Date Approved	June 2020
Compulsory/Optional	Compulsory	Date Modified	
Pre-requisites	None		
Co-requisites	None		

WELCOME

This course introduces students to advanced data programming and analysis techniques to process data sets and derive meaningful, statistically sound observations (e.g., summary statistics).

DESCRIPTION

Data Science is about extracting generalisable, useful and meaningful knowledge from data in a systematic manner. The course covers the emerging field of Data Science at breadth, honing on both programming and data analytics skills.

Students will learn to work with tensors (i.e., multi-dimensional arrays) and apply linear algebra transformations using, e.g., the `numpy` Python library; load, integrate and process structured and unstructured data from multiple sources using `pandas`; apply statistical and machine learning analysis algorithms using `scikit-learn`; and visualise results using `matplotlib`.

AIMS

The aim of the course is to set strong foundations in modern data science technologies and methods. More specifically, the course aims to:

- Develop familiarity with interactive and integrated development environments in Python
- Understand arrays and vectorised computation

- Understand data set representations of data transformations
- Develop ability to load and clean data sets, conduct descriptive, meaningful statistics and visualise results

LEARNING OUTCOMES

On successful completion of the course, students should be able to:

KNOWLEDGE AND UNDERSTANDING

K1a   show familiarity and knowledge of basic data science concepts

K2a   show knowledge of feasible operations on data and transformation on data

K3a   show knowledge and understanding of plotting and visualising data

K4a   demonstrate knowledge of applying techniques to a data case-study

SUBJECT-SPECIFIC SKILLS

S1a   engage in the theory behind the concepts taught in the class

S2a   apply the data transformation techniques in an appropriate manner to the chosen dataset

S3a   identify the correct choice of appropriate data transformation techniques

TRANSFERABLE SKILLS

T1a   work independently, effectively, and to deadlines

T2a   identify, transform, evaluate and plot accordingly from the dataset

T3a   produce clear and concise and well documented code

LEARNING AND TEACHING PLAN

Teaching and learning strategies for this course include:

- 10 × 1.5 hours of full-cohort lectures
- 10 × 1.5 hours of lab-based tutorials
- 2 office hours per teaching week

Course information and supplementary materials are available on the College's Virtual Learning Environment (VLE).

At the end of Hilary term, students will attend *Collections*, formal meetings where they receive comprehensive and collated feedback about their performance over the term.

Students are required to attend and participate in all the formal and timetabled sessions for this course. Students are also expected to manage their directed learning and independent study in support of the course.

FEEDBACK

Students receive feedback in a variety of ways:

- Written, including via email correspondence[1]

- Oral – during office hours, on an *ad hoc* basis, and during lectures and lab sessions

- *Collections* (see above)

Feedback is provided on both formative and summative assignments (normally, within a week of submission for formative and within two weeks for summative assignments) and it is made available on the College's VLE.

TEACHING SCHEDULE

Hilary Term		
Week	**Topic**	**Reading[2]**
1	Program design in Python - Introduction - Expressions - Objects - Control flow statements – conditionals & iterations - Data structures – lists, tuples & dictionaries - Program structure – functions & packages	Handbook: - §4.0-4.11 Scratch book: - Ch. 1 - Ch. 2 - Ch. 3
2	Data representations - Vectors, arrays & tensors - Algebraic operations - Pandas – series & data frames	Handbook: - Ch. 2 - §3.0 - §3.1 - Ch. 4 Scratch book: - Ch. 4
3	Data analytics - Relational algebra – selection, projection, join & aggregation operators - Data streams - MapReduce (data parallelism)	Handbook: - §3.2 - §3.3 - §§3.5-3.9 - §3.11 - §3.12 - Ch. 4

---

[1] **Email policy.** You are welcome to reach me, the course leader, via email and I aim to respond in a timely manner. Please ensure your email subject starts with "FODS:", followed by your subject. This way, I can identify and group your messages appropriately.

[2] "Handbook" refers to the book *Python Data Science Handbook*. "Scratch book" refers to the book *Data Science from Scratch*. See ESSENTIAL READING section.

Hilary Term		
Week	**Topic**	**Reading[2]**
		Scratch book: • Ch. 25
4	Data-driven applications (A and Ω) • Data collection • Web scraping • Data storage & retrieval • Data pre-processing • Data visualisation	Handbook: • §3.4 • §3.10 • Ch. 4 Scratch book: • Ch. 9
5	Statistics • Random variables • Probability distributions • Gaussian distribution & Mixtures of Gaussians • t-tests & Analysis of Variance (ANOVA) • Degrees of freedom	Scratch book: • Ch. 5 • Ch. 6 • Ch. 7
6	Regression • Curve fitting • Linear regression • Regularisation • Support Vector Machines • Clustering – k-nearest neighbours & k-means clustering • Principal Component Analysis	Handbook: • §§5.0-5.2 • §5.6 • §5.7 • §5.9 • §5.11 Scratch book: • Ch. 12 • Ch. 14 • Ch. 20
7	*Reading Week*	–
8	Classification • Bayesian models • Decision trees & random forests • Logistic regression • Neural networks	Handbook: • §5.5 • §5.8 Scratch book: • Ch. 13 • Ch. 17 • Ch. 18
9	Machine learning applications (A to Ω) • Feature engineering & selection • Model selection & model validation • Confusion matrix • Hyper-parameter tuning • Classification reports	Handbook: • §5.3 • §5.4
10	Text Analytics • Strings – text as data • Text pre-processing	Handbook: • §3.10

Hilary Term		
Week	**Topic**	**Reading²**
	• Sentiment analysis	Scratch book: • Ch. 21
11	Selected Topics in Data Science • Graph analytics • Practical ethical considerations	
12	*Collections*	–

Please refer to your CELCAT timetable for exact dates and times of lectures, seminars and tutorials.

ASSESSMENT

Please refer to the "Formative and Summative Assessment Planner" for the **submission dates** of your formative and summative assignments. All assignment briefs can be found on the on the course page on the College's VLE.

FORMATIVE

There is formative work every teaching week. Students will be given a mark for all formative work – not necessarily a precise numerical mark, but at least a qualitative indication of where a piece of work lies. The mark will be communicated to students at the time feedback is given, for it constitutes an element of that feedback:

Qualitative mark	Numerical range
First Class (1st)	72 and above
Second Class (First Division) (2:1)	62-68
Second Class (Second Division) (2:2)	52-58
Third Class (3rd)	42-48

SUMMATIVE

Students will be formally assessed with **two, equally weighted assignments**:

No.	Assignment Type	Weight (%)	Details	Length
AE1	Coding assignment	50	Online	Code and ~2,500 words
AE2	Coding assignment	50	Online	Code and ~2,500 words

READING
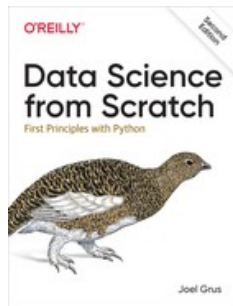
REQUIRED READING

Python Data Science Handbook

By Jake VanderPlas

2016

O'Reilly Media, Inc.

ISBN: 9781491912058

The book is available online on GitHub. It is also available for download from O'Reilly via Northeastern's University Library.

Data Science from Scratch, 2nd Edition

By Joel Grus

2019

O'Reilly Media, Inc.

ISBN: 9781492041139

The book is available online from O'Reilly via Northeastern's University Library (link).

FURTHER READING
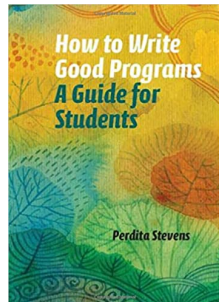
How to Write Good Programs: A Guide for Students

By Perdita Stevens

2020

Cambridge University Press

ISBN: 9781108789875

The book is available online for download from the publisher via Northeastern's University Library (link).

CHALLENGING READING

MapReduce: simplified data processing on large clusters • J. Dean and S. Ghemawat • Communications of the ACM • January 2008 • DOI

A relational model of data for large shared data banks • E. F. Codd • Communications of the ACM • June 1970 • DOI

SciPy Lecture Notes – One document to learn numerics, science and data with Python • link

Finding the frequent items in streams of data • G. Cormode and M. Hadjieleftheriou • Communications of the ACM • October 2009 • link

APPENDIX A: COMMON ASSESSMENT MARKING SCHEME

Class				
1st	Upper 2nd	Lower 2nd	3rd	Fail
100	68	58	48	35
90	65	55	45	20
85	62	52	42	5
80				0
75				
72				

APPENDIX B: GENERIC GRADE CRITERIA – LEVEL 4

KNOWLEDGE AND UNDERSTANDING						
100/90/85/80	**75/72**	**68/65/62**	**58/55/52**	**48/45/42**	**35**	**20/5/0**
Exceptional breadth and depth for work at this level.	Accurate and coherent in breadth, with depth in many areas.	Accurate in breadth, with depth in several areas.	Accurate, with depth in some aspects.	Largely accurate across most areas, with limited depth.	Inaccuracies/omissions in some areas, depth limited.	Substantial inaccuracies, omissions, irrelevancies.
Excellent understanding of concepts/theories (some of them abstract) and/or current practice, and several of their applications and implications.	Thorough understanding of concepts and theories (some of them abstract) and/or current practice, and some of their implications and applications.	Clear understanding of concepts and theories (some of them abstract) and/or practice and some of their implications and applications.	Satisfactory understanding of the relevant concepts, theories and/or practice; Shows some ability to deal with unfamiliar and abstract ideas.	Adequate understanding of the main concepts, theories, and/or practice; Engagement with unfamiliar/ abstract ideas or implications and applications is slight.	Occasional errors in understanding of main concepts, theories and/or practice; Struggles to engage with unfamiliar/ abstract ideas and complexities.	Substantial errors in understanding of concepts, theories and/or practice, or none.

SUBJECT SPECIFIC						
100/90/85/80	**75/72**	**68/65/62**	**58/55/52**	**48/45/42**	**35**	**20/5/0**
Selects and applies appropriate methods to address/solve complex and often unfamiliar and unpredictable problems.	Applies appropriate methods to address/solve complex issues/problems, some unfamiliar/ Unpredictable.	Uses appropriate given methods to address complex issues/ problems, some unfamiliar /unpredictable.	Uses given methods to analyse issues/ problems, some unfamiliar /unpredictable and complex.	Analysis using given methods is adequate.	Superficial analysis.	Analysis absent or with significant errors/ Omissions.
Exceptional judgement in selection, analysis and evaluation of information and application of learning to different contexts.	Exercises judgement in selection, analysis and evaluation of information and application of learning to a different context.	Exercises judgement in selection and analysis of information, with some evaluation, and application of learning in a different context.	Satisfactory selection and analysis of information, with little evaluation; Applies some aspect of learning in a different context.	Limited ability to apply learning to complex, unfamiliar or unpredictable contexts or issues.	Some failure to apply learning complex, unfamiliar or unpredictable issues/contexts.	Fails to apply learning.
Excellent investigative skills generate well-founded and evidenced conclusions /practical solutions.	Thorough investigation generates well-founded conclusions/practical solutions.	Investigation generates well-founded conclusions /practical solutions.	Investigation generates some conclusions/ practical solutions.	Tendency to description and reliance on familiar/ given methods and approaches.	Overly descriptive and reliant on familiar/given material or approaches.	Descriptive and heavily reliant on very restricted range of given/familiar material and approaches, poorly understood.
Explores and evaluates	Explores and deploys information	Locates and organises a wide	Locates and organises a	Locates and organises an	Range of information limited	Range of information

SUBJECT SPECIFIC						
100/90/85/80	**75/72**	**68/65/62**	**58/55/52**	**48/45/42**	**35**	**20/5/0**
information/ideas from a wide range of sources (may include primary sources).	from a wide range of mostly secondary sources.	range of information/evidence.	satisfactory range of information/evidence, some of it beyond the given/familiar.	acceptable range of information/evidence mostly from given/ familiar secondary sources.	to the familiar/ given with some errors in organisation.	inadequate and disorganised.
Competence in all the required specialised practical, technical, creative, scholarly or work-related skills. exceeds expectations for this level.	Competence in all the required specialised practical, technical, creative, scholarly or work-related skills, exceeds expectations for this level in some aspects.	Competently uses all the required specialised practical, technical, creative, scholarly or work-related skills, with indications of more developed ability in some areas.	Competently uses all of the required specialised practical, technical, creative, scholarly or work-related skills, with more developed capability in at least one area.	Use of all the required specialised practical, technical, creative, scholarly, or work-related skills is adequate.	Use of some of the required specialised practical, technical, creative, scholarly or work-related skills is inadequate.	Inadequate use of many/all of the required specialised practical, technical, creative, scholarly or work related skills.

TRANSFERABLE SKILLS						
100/90/85/80	**75/72**	**68/65/62**	**58/55/52**	**48/45/42**	**35**	**20/5/0**
Excellent presentation and organisation of work and lucid communication in all contexts.	Excellent presentation and organisation of work and lucid communication in most contexts.	Presentation and organisation of work appropriate to context and purpose, communication clear.	Satisfactory organisation and presentation of work, communications mostly appropriate to the context/purpose.	Organisation and presentation of work and communications adequate in most contexts; some mistakes/irrelevancies.	Elements of disorganisation/ poor presentation/ poor or inappropriate communication or expression.	Work is disorganised, poorly presented with poor inappropriate communication and expression.
Exemplary referencing/citation.	Extensive, accurate referencing/citation.	Referencing consistent and accurate.	Referencing mostly consistent/accurate.	Some errors in referencing.	Errors/omissions in referencing, or none.	Substantial errors in referencing, or none.
Work demonstrates independence and initiative beyond level expectations, setting objectives and taking responsibility for outcomes.	Work demonstrates independence and some initiative in setting objectives and taking responsibility for outcomes.	Work demonstrates independence in setting some objectives beyond those given and taking responsibility for outcomes.	Work demonstrates satisfactory independence in addressing objectives and taking responsibility for outcomes.	Work demonstrates adequate independence in taking responsibility for outcomes.	Work demonstrates insufficient independence in attempting to address given objectives and taking responsibility for outcomes.	Work fails to address objectives and take responsibility for outcomes.
Evidences developed team-working and indications of leadership ability.	Evidences developed team-working skills.	Evidences a high level of team-working skills.	Evidences team-working and basic leadership skills.	Tendency to rely on support/direction from others.	Over-reliance on support/direction from others.	Fails to engage in /shows deficiencies in team working.
Critical reflection/self-evaluation exceptional	Reflection and self-evaluation often	Reflection generates a	Satisfactory reflection with	Limited reflection	Minimal reflection	Reflection inadequate/absent

TRANSFERABLE SKILLS						
100/90/85/80	**75/72**	**68/65/62**	**58/55/52**	**48/45/42**	**35**	**20/5/0**
for this level.	critical and insightful.	number of critical insights.	some insights.	with few insights	lacks insight.	with no insight.