Liverpool John Moores University

Title:	Efficient algorithms for complex data sets
Status:	Definitive
Code:	**7022DATSCI** (125177)
Version Start Date:	01-08-2021
Owning School/Faculty:	Astrophysics Research Institute
Teaching School/Faculty:	Astrophysics Research Institute

Team	Leader
Ivo Siekmann	Y
Paolo Mazzali	

Academic Level:	FHEQ7	**Credit Value:**	20	**Total Delivered Hours:**	60
Total Learning Hours:	200	**Private Study:**	140		

Delivery Options
Course typically offered: Semester 2

Component	Contact Hours
Lecture	15
Seminar	10
Workshop	35

Grading Basis: 50 %

Assessment Details

Category	Short Description	Description	Weighting (%)	Exam Duration
Report	AS1	Report based on methods for analysing complex data sets	60	
Test	AS2	In-class Test	40	2

Aims

The module provides an introduction to algorithms for the analysis of complex data sets. This includes high-dimensional data sets which require pre-processing using efficient dimensionality reduction and visualisation techniques. Other complex data sets come in the form of large graphs (networks) or have been generated by

complex underlying processes. Especially for the latter class of data sets the module aims to develop the key skill of working with experts from other domains.

Learning Outcomes

After completing the module the student should be able to:

1	Synthesize a combination of various machine learning and pre-processing techniques to explore large data sets.
2	Apply efficient dimensionality reduction and distributed & parallel data processing.
3	Evaluate different incremental machine learning approaches.

Learning Outcomes of Assessments

The assessment item list is assessed via the learning outcomes listed:

Report on big data analysis	1	2	3
In-class Test	1	3	

Outline Syllabus

1. Key Linear Algebra Techniques (vectors, matrices, numerical calculation of eigenvalues and eigenvectors)
2. Dimensionality Reduction (statistical methods and random projections)
3. Network Analysis
4. Monte Carlo methods for analysis and simulation of complex systems (e.g. Supernova atmosphere).
5. Computational Bayesian statistics and Markov Chain Monte Carlo (MCMC)
6. Practical implementation implementation of statistical and machine learning techniques using distributed and parallel data processing.

Learning Activities

Lectures
Seminar discussions
Directed Reading
Computer based exercises

Notes

.