

Summary Information

Module Code	7022DATSCI
Formal Module Title	Efficient algorithms for complex data sets
Owning School	Astrophysics Research Institute
Career	Postgraduate Taught
Credits	20
Academic level	FHEQ Level 7
Grading Schema	50

Module Contacts

Module Leader

Contact Name	Applies to all offerings	Offerings
Ivo Siekmann	Yes	N/A

Module Team Member

Contact Name	Applies to all offerings	Offerings
Paolo Mazzali	Yes	N/A

Partner Module Team

Contact Name	Applies to all offerings	Offerings
--------------	--------------------------	-----------

Teaching Responsibility

LJMU Schools involved in Delivery
Astrophysics Research Institute
Computer Science and Mathematics

Learning Methods

Learning Method Type	Hours
Lecture	15
Seminar	10
Workshop	35

Module Offering(s)

Offering Code	Location	Start Month	Duration
JAN-CTY	CTY	January	12 Weeks

Aims and Outcomes

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:

Code	Description
MLO1	Synthesize a combination of various machine learning and pre-processing techniques to explore large data sets.
MLO2	Apply efficient dimensionality reduction and distributed & parallel data processing.
MLO3	Evaluate different incremental machine learning approaches.

Module Content

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 Analysis4. 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 of statistical and machine learning techniques using distributed and parallel data processing.

Module Overview

This module aims to develop skills in big data analysis, including techniques of dimensionality reduction and the application of statistical and machine learning models. In addition it aims to develop the key skill of working with experts from other domains.

Additional Information

Assessments

Assignment Category	Assessment Name	Weight	Exam/Test Length (hours)	Learning Outcome Mapping
Report	Report on big data analysis	60	0	MLO3, MLO1, MLO2
Exam	In-class Test	40	2	MLO3, MLO1