

Multivariate Analysis and Data Mining

Module Information

2022.01, Approved

Summary Information

Module Code	6106STATS
Formal Module Title	Multivariate Analysis and Data Mining
Owning School	Computer Science and Mathematics
Career	Undergraduate
Credits	20
Academic level	FHEQ Level 6
Grading Schema	40

Teaching Responsibility

LJMU Schools involved in Delivery
Computer Science and Mathematics

Learning Methods

Learning Method Type	Hours
Lecture	49
Practical	6

Module Offering(s)

Display Name	Location	Start Month	Duration Number Duration Unit
JAN-CTY	CTY	January	12 Weeks

Aims and Outcomes

Aims	To enable the student to explore the structure of multidimensional data sets.To introduce the student to inferential procedures using multivariate data.To provide the student with familiarity with linear and flexible methods for regression and classification.
------	---

After completing the module the student should be able to:

Learning Outcomes

Code	Number	Description
MLO1	1	Carry out an exploratory numerical and graphical analysis of a set of multivariate data.
MLO2	2	Recognize situations in which a multivariate approach is required and carry out the appropriate inferential procedures.
MLO3	3	Classify and cluster future multivariate observations into one of a number of known populations.
MLO4	4	Apply linear methods and neural network algorithms.
MLO5	5	Write a brief report.

Module Content

Outline Syllabus	Graphical display and numerical summary of multivariate data. Investigation of the dependence among variables. Discrimination and prediction. Error rate estimation. Hypothesis construction and testing. Use of simultaneous confidence intervals. Principal Components Analysis. Use of R for data exploration, parameter estimation and significance testing. Stages of the data analysis using data mining, including how to create and evaluate models. Neural networks, including backpropagation, gradient descent, the momentum method. Clustering with k-means and k-medoids methods. Blind Source Separation, including Independent Component Analysis and Non-negative Matrix Factorisation. Self-Organising Maps. Introduction to Big Data
Module Overview	
Additional Information	Module Overview This final year module advances beyond univariate modelling with statistical methods, to the analysis of multivariate data. There are two generally accepted areas of statistics: exploratory and inferential. How does this Module relate to the Programme overall? The module begins with exploratory numerical and graphical methods, moving on to predictive inference with linear-in-the-parameter mainstream statistical methods. During the second semester, the module serves to introduce the use of non-linear methods. This begins with flexible extensions of generalised linear models, which are also feature in the machine learning literature under the name of neural networks. This reflects a link between statistics, computational learning and biological systems, whereby inference models are estimated by adjusting link-weights that control communication of information between simple neural cells. This leads to the relatively new field of data mining, which in essence comprises linear and non-linear exploration of data, through visualisation, clustering and inferential modelling.

Assessments

Assignment Category	Assessment Name	Weight	Exam/Test Length (hours)	Module Learning Outcome Mapping
Centralised Exam	Examination	60	2	MLO1, MLO3, MLO4
Portfolio	Coursework portfolio	40	0	MLO2, MLO5, MLO4

Module Contacts

Module Leader

Contact Name	Applies to all offerings	Offerings
Sandra Ortega Martorell	Yes	N/A

Partner Module Team

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