ST340-15 Fundamentals of Machine Learning

26/27

Department
Statistics
Level
Undergraduate Level 3
Module leader
Paul Jenkins
Credit value
15
Module duration
10 weeks
Assessment
Multiple
Study location
University of Warwick main campus, Coventry

Description

Introductory description

This module introduces students to the theoretical foundations and key algorithms of modern machine learning.

Pre-requisites

- Statistics students. ST231 Linear Statistical Modelling with R .
- Non-statistics students. ST240 Linear Statistical Modelling or ST351 Linear Statistical Modelling (For Finalists).

This module is available for students on a course where it is a listed option and as an Unusual Option to students who have the required background as covered by the pre-requisite modules.

Pre-registration. There is a cap on student numbers for this module and pre-registration is essential. Information about prioritisation and the pre-registration form can be found at http://go.warwick.ac.uk/ST340

Leads to: ST349 Machine Learning Frameworks, ST420 Statistical Learning and Big Data.

Module web page

Module aims

1. Develop students' understanding of the theoretical foundations of modern machine learning algorithms.
2. Introduce key algorithms for both supervised and unsupervised learning, with simple implementations in R.
3. Provide relevant background to prepare students for further advanced modules in machine learning and for a career in data science.

Outline syllabus

This is an indicative module outline only to give an indication of the sort of topics that may be covered. Actual sessions held may differ.

The module will cover many contemporary data analysis problems and algorithms.

1. Fundamentals of statistical learning: Supervised vs unsupervised, regression vs classification, over- vs under-fitting, curse of dimensionality, training vs testing, validation and cross-validation, regularisation.
2. Supervised learning methods. An indicative list is: logistic regression, k-nearest neighbours, discriminant analysis, Gaussian processes, artificial neural networks; exact details may vary from year to year.
3. Unsupervised learning methods. An indicative list is: K-means clustering, dimension reduction such as principal component analysis; exact details may vary from year to year.

Learning outcomes

By the end of the module, students should be able to:

- Evaluate data analysis problems and select appropriate supervised and unsupervised learning algorithms.
- Critically evaluate the strengths and weaknesses of machine learning algorithms.
- Implement data analysis algorithms.
- Interpret the output of various algorithms when applied to data sets.

Indicative reading list

Reading lists can be found in Talis

Specific reading list for the module

Subject specific skills

- Demonstrate facility with rigorousdata science methods.

- Evaluate, select and apply appropriate techniques to a variety of situations.
- Demonstrate knowledge of and facility with data science concepts, both explicitly and by applying them to the solution of problems.
- Create structured and coherent arguments communicating them in written form.
- Construct logical mathematical arguments with clear identification of assumptions and conclusions.
- Reason critically, carefully, and logically.

Transferable skills

- Problem solving: Use rational and logical reasoning to deduce appropriate and well-reasoned conclusions. Retain an open mind, optimistic of finding solutions, thinking laterally and creatively to look beyond the obvious. Know how to learn from failure.
- Self awareness: Reflect on learning, seeking feedback on and evaluating personal practices, strengths and opportunities for personal growth.
- Communication: Present arguments, knowledge and ideas, in a range of formats.
- Professionalism: Prepared to operate autonomously. Aware of how to be efficient and resilient. Manage priorities and time. Self-motivated, setting and achieving goals, prioritising tasks.

Study

Study time

Type	Required
Lectures	25 sessions of 1 hour (17%)
Practical classes	5 sessions of 1 hour (3%)
Private study	88 hours (59%)
Assessment	32 hours (21%)
Total	150 hours

Private study description

Weekly revision of lecture notes and materials, wider reading, practice exercises and preparing for coursework and examination.

Other activity description

Revision support.

Costs

No further costs have been identified for this module.

Assessment

You do not need to pass all assessment components to pass the module.

Assessment group D

	Weighting	Study time	Eligible for self-certification
Assignment 1	10%	15 hours	No

You will use R to implement and run algorithms on large datasets in response to a set of questions. You will present, discuss and evaluate the results. The study time noted below refers to the amount of time in hours that a well-prepared student who has attended lectures and carried out an appropriate amount of independent study on the material could expect to spend on this assignment. Your ST340 Assignment should not exceed 15 pages in length.

Assignment 2	10%	15 hours	No

You will analyse algorithms. You will use R to implement algorithms in response to a set of questions. You will present, discuss and evaluate the results. The study time noted below refers to the amount of time in hours that a well-prepared student who has attended lectures and carried out an appropriate amount of independent study on the material could expect to spend on this assignment. Your ST340 Assignment should not exceed 15 pages in length.

Centrally-timetabled examination (On-campus)	80%	2 hours	No

The examination paper will contain four questions, of which the best marks of THREE questions will be used to calculate your grade.

The study time noted refers to the length of exam in hours.

- Answerbook Pink (12 page)

Assessment group R5

	Weighting	Study time	Eligible for self-certification
In-person Examination - Resit	100%		No

The examination paper will contain four questions, of which the best marks of THREE questions

	Weighting	Study time	Eligible for self-certification

will be used to calculate your grade.

- Answerbook Pink (12 page)

Feedback on assessment

Assignments are marked and given feedback online within 20 working days of the submission deadline.

Solutions and cohort level feedback will be provided for the examination

Students are actively encouraged to make use of office hours to build up their understanding, and to view all their interactions with lecturers and class tutors as feedback.

Past exam papers for ST340

Availability

Courses

This module is Optional for:

- USTA-G300 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics
  - Year 3 of G300 Mathematics, Operational Research, Statistics and Economics
  - Year 4 of G300 Mathematics, Operational Research, Statistics and Economics

This module is Option list A for:

- Year 4 of USTA-G1G3 Undergraduate Mathematics and Statistics (BSc MMathStat)
- Year 5 of USTA-G1G4 Undergraduate Mathematics and Statistics (BSc MMathStat) (with Intercalated Year)
- Year 3 of USTA-GG14 Undergraduate Mathematics and Statistics (BSc)
- Year 4 of USTA-GG17 Undergraduate Mathematics and Statistics (with Intercalated Year)
- Year 3 of USTA-Y602 Undergraduate Mathematics,Operational Research,Statistics and Economics
- Year 4 of USTA-Y603 Undergraduate Mathematics,Operational Research,Statistics,Economics (with Intercalated Year)

This module is Option list B for:

- Year 3 of USTA-G302 Undergraduate Data Science
- Year 3 of USTA-G304 Undergraduate Data Science (MSci)
- Year 4 of USTA-G303 Undergraduate Data Science (with Intercalated Year)

- UMAA-G105 Undergraduate Master of Mathematics (with Intercalated Year)
    - Year 4 of G105 Mathematics (MMath) with Intercalated Year
    - Year 5 of G105 Mathematics (MMath) with Intercalated Year
- Year 3 of UMAA-G100 Undergraduate Mathematics (BSc)
- UMAA-G103 Undergraduate Mathematics (MMath)
    - Year 3 of G100 Mathematics
    - Year 3 of G103 Mathematics (MMath)
    - Year 4 of G103 Mathematics (MMath)
- Year 4 of UMAA-G107 Undergraduate Mathematics (MMath) with Study Abroad
- Year 3 of USTA-G1G3 Undergraduate Mathematics and Statistics (BSc MMathStat)
- Year 4 of USTA-G1G4 Undergraduate Mathematics and Statistics (BSc MMathStat) (with Intercalated Year)
- Year 4 of UMAA-G101 Undergraduate Mathematics with Intercalated Year

This module is Option list D for:

- Year 5 of USTA-G301 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics (with Intercalated

This module is Option list E for:

- Year 5 of USTA-G301 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics (with Intercalated

This module is Option list F for:

- USTA-G301 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics (with Intercalated
    - Year 3 of G30H Master of Maths, Op.Res, Stats & Economics (Statistics with Mathematics Stream)
    - Year 4 of G30H Master of Maths, Op.Res, Stats & Economics (Statistics with Mathematics Stream)

This module is Option list G for:

- USTA-G300 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics
    - Year 3 of G300 Mathematics, Operational Research, Statistics and Economics
    - Year 4 of G300 Mathematics, Operational Research, Statistics and Economics