ST349-15 Machine Learning Frameworks

26/27

Department
Statistics
Level
Undergraduate Level 3
Module leader
Wenkai Xu
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 contemporary practice of machine learning and deep learning. The module takes a hands-on approach where concepts are introduced using the various machine learning and deep learning frameworks. The module is theoretical and students will be expected to have a level of programming ability to implement this theory.

Prerequisites. This module assumes you have studied and completed ST231 Linear Statistical Modelling with R or ST240 Linear Statistical Modelling/ST351 Linear Statistical Modelling (For Finalists).

Preregistration. 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/ST349

Recommended. It is recommended that you have studied and completed ST236 Python for Data Analytics Tasks and ST340 Programming for Data Science/Fundamentals of Machine Learning. You will need to have sufficient programming proficiency to implement programming tools fluently.

Module web page

Module aims

The module aims to develop knowledge of:

1. Machine learning and deep learning concepts, tasks, models, and workflows.
2. Machine learning frameworks.
3. Deep learning frameworks.
4. Communicating outputs from machine learning and deep learning frameworks to literate audiences.

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.

This module covers the following topics.

1. Modes of statistical learning / common machine learning tasks. Supervised learning; unsupervised learning; semi-supervised learning.
2. Machine learning frameworks. Working with machine learning frameworks in Python (Scikit-Learn, Pytorch, graphlearning).
3. Unsupervised learning / dimensionality reduction algorithms: for example, principal component analysis, singular value decomposition, k-means clustering, spectral clustering, mixture models, DBSCAN, OPTICS, generative models.
4. Semi-supervised learning algorithms: for example, nearest neighbour, Laplace learning, Poisson learning, kernel-based method.

Learning outcomes

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

- Appraise and decide on the appropriate type of statistical learning required to address a real-world challenge.
- Create data-analytic pipelines and workflows for the various modes of statistical learning.
- Implement common machine learning and deep learning frameworks in Python.
- Collaborate and disseminate fully documented Python code with reproducible outputs.
- Interpret and communicate the outputs of a completed analysis to a range of audiences.

Indicative reading list

Reading lists can be found in Talis

Interdisciplinary

This module requires students to develop a balanced facility of familiarity with machine learning frameworks and data-analytic skills for solving real-world problems across disciplines.

Subject specific skills

1. Demonstrate advanced facility with data handling and analysis methods in R and Python.
2. Create readable, valid, reliable, reproducible and well-documented code.
3. Appraise problems, abstracting their essential information to make judgements on the appropriate concepts to facilitate their solution.
4. Demonstrate programming skills and knowledge of programming concepts, both explicitly and by
applying them to the solution of real-world problems.

Transferable skills

1. Problem-solving skills: The module requires students to solve problems and present their conclusions as logical and coherent arguments.
2. Written communication skills: Students complete written assessments that require precise and unambiguous communication in the manner and style expected in mathematical sciences.
3. Verbal communication skills: Students are encouraged to discuss and debate formative assessment and lecture material within small-group tutorial sessions. Students can continually discuss specific aspects of the module with the module leader. This is facilitated by statistics staff office hours.
4. Team working and working effectively with others: Students are encouraged to discuss and debate formative assessment and lecture material within small-group tutorial sessions.
5. Professionalism: Students work autonomously by developing and sustaining effective approaches to learning, including time management, organisation, flexibility, creativity, collaboratively and intellectual integrity.

Study

Study time

Type	Required
Lectures	20 sessions of 1 hour (13%)
Practical classes	10 sessions of 1 hour (7%)
Private study	76 hours (51%)
Total	150 hours

Type	Required
Assessment	44 hours (29%)
Total	150 hours

Private study description

Weekly revision of lecture notes and materials, wider reading and practice exercises, working on problem sets and preparing for 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 D1

	Weighting	Study time	Eligible for self-certification
Individual Assignment	15%	10 hours	No
An individual assignment to professional standards, presenting the analysis, interpretation and conclusion of the task set. All code must be version-controlled, well-documented and reproducible. Submitted code will be part of the assignment's appendix and will not count toward the page limit. This assignment should not exceed 8 pages in length.			
Group Assignment	15%	10 hours	No
A formal group report, to professional standards, presenting the analysis, interpretation and conclusion of the task set. All code must be version-controlled, well-documented and reproducible. The target audience are decision-makers who do not necessarily have advanced statistical training. Submitted code will be part of the report's appendix and will not count toward the page limit. This report should not exceed 8 pages in length.			
Centrally-timetabled examination (On-campus)	70%	24 hours	No
The examination paper will contain four questions, of which the best marks of THREE questions will be used to calculate your grade.			

	Weighting	Study time	Eligible for self-certification

- Answerbook Pink (12 page)

Assessment group R2

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

The examination paper will contain four questions, of which the best marks of THREE questions 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. Where
appropriate, model solutions will be provided.

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 ST349

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 3 of USTA-G300 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics
- Year 4 of USTA-G1G3 Undergraduate Mathematics and Statistics (BSc MMathStat)

- Year 3 of USTA-GG14 Undergraduate Mathematics and Statistics (BSc)
- Year 4 of USTA-G140 Undergraduate Mathematics and Statistics (MMathStat)
- 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)
- 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 3 of USTA-G140 Undergraduate Mathematics and Statistics (MMathStat)
- Year 4 of UMAA-G101 Undergraduate Mathematics with Intercalated Year

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