1. **Title of the module**

MAST6053 - Statistical Learning for Data Scientists

MAST7053 – Statistical Learning for Data Scientists

1. **School or partner institution which will be responsible for management of the module**

School of Mathematics, Statistics and Actuarial Science

1. **The level of the module (e.g. Level 4, Level 5, Level 6 or Level 7)**

MAST6053 Level 6

MAST7053 Level 7

1. **The number of credits and the ECTS value which the module represents**

15 credits (7.5 ECTS)

1. **Which term(s) the module is to be taught in (or other teaching pattern)**

Spring

1. **Prerequisite and co-requisite modules**

**Level 6:**

Pre-requisite: MAST4009 (Probability); MAST4011 (Statistics); MAST5007 (Mathematical Statistics) or MAST5001 (Applied Statistical Modelling 1)

Co-requisite: None

**Level 7**:

Pre-requisite: Students are expected to have studied material equivalent to that covered in the modules above.

Co-requisite: None

1. **The courses of study to which the module contributes**

For the level 6 module, BSc Mathematics, BSc Mathematics and Statistics, BSc Financial Mathematics, BA Mathematics and Accounting and Finance (including courses with a Year in Industry), BSc Mathematics with a Foundation Year, MMath Mathematics.

For the level 7 module, MSc in Statistical Data Science, also with an Industrial Placement,

1. **The intended subject specific learning outcomes.**

**On successfully completing the level 6 module students will be able to:**

1. demonstrate systematic understanding of key aspects of multivariate and computational statistics and machine learning;
2. demonstrate the capability to deploy established approaches accurately to analyse and solve problems using a reasonable level of skill in calculation and manipulation of the material in the following areas: multivariate statistics, clustering (e.g., mixture modelling), classification, machine learning methods (e.g. graphical models) , maximum likelihood estimation, the EM algorithm and simulation methods
3. apply key aspects of multivariate and computational statistics and machine learning in well-defined contexts, showing judgement in the selection and application of tools and techniques;
4. show judgement in the application of R.

**On successfully completing the level 7 module students will be able to:**

1. demonstrate systematic understanding of multivariate and computational statistics and machine learning;
2. demonstrate the capability to solve complex problems using a very good level of skill in calculation and manipulation of the material in the following areas: multivariate statistics, clustering (e.g., mixture modelling), classification, machine learning methods such as graphical models , maximum likelihood estimation, the EM algorithm and simulation methods
3. apply a range of concepts and principles in multivariate and computational statistics and machine learning in loosely defined contexts, showing good judgment in the selection and application of tools and techniques;
4. make effective and well-considered use of R.
5. **The intended generic learning outcomes.**

**On successfully completing the level 6 module students will be able to:**

1. manage their own learning and make use of appropriate resources;
2. understand logical arguments, identifying the assumptions made and the conclusions drawn;
3. communicate straightforward arguments and conclusions reasonably accurately and clearly;
4. manage their time and use their organisational skills to plan and implement efficient and effective modes of working;
5. solve problems relating to qualitative and quantitative information;
6. make competent use of information technology skills such as online resources (Moodle), internet communication;
7. communicate technical and non-technical material competently;
8. demonstrate an increased level of skill in numeracy and computation;
9. demonstrate the acquisition of the study skills needed for continuing professional development.

**On successfully completing the level 7 module students will be able to:**

1. work competently and independently, be aware of their own strengths and understand when help is needed;
2. demonstrate a high level of capability in developing and evaluating logical arguments;
3. communicate arguments confidently with the effective and accurate conveyance of conclusions;
4. manage their time and use their organisational skills to plan and implement efficient and effective modes of working;
5. solve problems relating to qualitative and quantitative information;
6. make effective use of information technology skills such as online resources (Moodle), internet communication;
7. communicate technical and non-technical material effectively;
8. demonstrate an increased level of skill in numeracy and computation;
9. demonstrate the acquisition of the study skills needed for continuing professional development.
10. **A synopsis of the curriculum**

Background material: multivariate normal distribution, inference from multivariate normal samples

Indicative module content:

* Principal component and factor analysis, latent variable model, clustering and classification methods
* Likelihood-based analysis such as maximum likelihood, EM algorithm, optimisation, confidence interval construction
* Simulation and sampling methods, bootstrap, permutation tests
* Model building including tests such as the Wald test
* R programming including real-world applications in areas such as biology, ecology, sociology and economics to data that does not always follow standard statistical models.

In addition, for level 7 students: advanced EM algorithm methods, advanced simulation methods, writing R programs for advanced methods and applications.

1. **Reading List (Indicative list, current at time of publication. Reading lists will be published annually)**

D. F. Morrision (1990). Multivariate Statistical Methods, McGraw-Hill Series in Probability and Statistics  
T. Hastie, R. Tibshirani and J. H. Friedman (2009). The Elements of Statistical Learning, Spring-Verlag.  
K. P. Murphy (2012). Machine Learning: A Probabilistic Perspective, MIT Press.

Morgan, B. J. T. (2009) Applied stochastic modelling, Chapman and Hall.

1. **Learning and Teaching methods**

Total contact hours: 36

Private study hours: 114

Total study hours: 150

1. **Assessment methods**

13.1 Main assessment methods

**Level 6**

Assessment 1 Exercises, requiring on average between 10 and 15 hours to complete 10%

Assessment 2 Exercises, requiring on average between 10 and 15 hours to complete 10%

At least one assessment will include the use of R.

Examination 2 hours 80%

The coursework mark alone will not be sufficient to demonstrate the student’s level of achievement on the module.

**Level 7**

Assessment 1 Exercises, requiring on average between 10 and 15 hours to complete 10%

Assessment 2 Exercises, requiring on average between 10 and 15 hours to complete 10%

At least one assessment will include the use of R.

Examination 2 hours 80%

The coursework mark alone will not be sufficient to demonstrate the student’s level of achievement on the module.

13.2 Reassessment methods

Like-for-like

1. **Map of Module Learning Outcomes (sections 8 & 9) to Learning and Teaching Methods (section12) and methods of Assessment (section 13)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Level 6 Module learning outcome** | 8.1 | 8.2 | 8.3 | 8.4 | 9.1 | 9.2 | 9.3 | 9.4 | 9.5 | 9.6 | 9.7 | 9.8 | 9.9 |
| **Learning/ teaching method** | | | | | | | | | | | | | |
| Private Study | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |
| Lectures/Exercise classes | **X** | **X** | **X** | **X** |  | **X** | **X** |  | **X** |  | **X** | **X** |  |
| Terminal classes | **X** | **X** | **X** | **X** |  | **X** | **X** |  | **X** | **X** | **X** | **X** |  |
| Revision classes | **X** | **X** | **X** |  |  | **X** | **X** |  | **X** |  | **X** | **X** |  |
| **Assessment method** | | | | | | | | | | | | | |
| Examination | **X** | **X** | **X** |  | **X** | **X** | **X** | **X** | **X** |  | **X** | **X** | **X** |
| Coursework | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Level 7 Module learning outcome** | 8.5 | 8.6 | 8.7 | 8.8 | 9.10 | 9.11 | 9.12 | 9.13 | 9.14 | 9.15 | 9.16 | 9.17 | 9.18 |
| **Learning/ teaching method** | | | | | | | | | | | | | |
| Private Study | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |
| Lectures/Exercise classes | **X** | **X** | **X** | **X** |  | **X** | **X** |  | **X** |  | **X** | **X** |  |
| Terminal classes | **X** | **X** | **X** | **X** | **X** | **X** | **X** |  | **X** | **X** | **X** | **X** | **X** |
| Revision classes | **X** | **X** | **X** |  |  | **X** | **X** |  | **X** |  | **X** | **X** |  |
| **Assessment method** | | | | | | | | | | | | | |
| Examination | **X** | **X** | **X** |  | **X** | **X** | **X** | **X** | **X** |  | **X** | **X** | **X** |
| Coursework | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |

1. **Inclusive module design**

The School recognises and has embedded the expectations of current equality legislation, by ensuring that the module is as accessible as possible by design. Additional alternative arrangements for students with Inclusive Learning Plans (ILPs)/declared disabilities will be made on an individual basis, in consultation with the relevant policies and support services.

The inclusive practices in the guidance (see Annex B Appendix A) have been considered in order to support all students in the following areas:

a) Accessible resources and curriculum

b) Learning, teaching and assessment methods

1. **Campus(es) or centre(s) where module will be delivered:** Canterbury
2. **Internationalisation**

Statistics as a branch of mathematics is an international language with techniques developed and refined by statisticians across the globe. Mastery of the subject-specific learning outcomes, 8.1 to 8.8, will equip students to apply the theories and techniques of this module in a wide range of international contexts. The module team is drawn from the School of Mathematics, Statistics and Actuarial Science, which includes many members of staff with international experience of teaching and research collaboration.

In compiling the reading list, consideration has been given to the range of texts that are available internationally and a selection of texts has been identified to complement the delivery of the material.

Examples with an international dimension are included in the module where appropriate.

The support SMSAS provides to its students is also internationally attuned given our international student body.

**DIVISIONAL USE ONLY**

**Revision record – all revisions must be recorded in the grid and full details of the change retained in the appropriate committee records.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date approved | Major/minor revision | Start date of delivery of revised version | Section revised | Impacts PLOs (Q6&7 cover sheet) |
| 1 July 2021 | New Module | 2022/23 |  |  |
|  |  |  |  |  |