School of Informatics Course Proposal Form (version: May 2021)

Please see Page 2 for instructions on which parts of this form to complete, whom to consult with to avoid unnecessary effort, and where to send the completed form.

Proposer(s): Hao Tang and Chris Williams Date: 10 A9 August 2021

Cover page: Basic permanent course information

Unless otherwise noted, items in this section are entered into EUCLID and **cannot** be changed without creating an entirely new course.

Course Name	Machine Learning	
Is this an EPCC course?	<u>x</u> No (default) (If you don't know what EPCC is, this is the right choice.) Yes (If so, leave Course Acronym blank, to be filled in by ITO as EPCC/ <number> for Theon and our Sortable List.)</number>	
Course Acronym (used only School-internally)	MLG	
SCQF Credit Level and Normal Year Taken	MLG Standard options for Informatics courses: Level 8/Year 1 Level 8/Year 2 X. Level 10/Year 3 (also available in Year 4). [In practice, most level 10 courses have many students in both UG3 and UG4. MSc students may take up to 20 credits at Level 10.] Level 11/Year 4 (also available in Year 5 and MSc). [These courses are listed as options in both UG and MSc DPTs.] Level 11/PG (also available in Year 5). [These courses are normally for MSc and UG5 students. They are not explicitly listed in UG4 DPTs, but UG4 students can take limited credits of them.] Level 11/PG (only). [These courses are not available to UG4 or UG5 students. Examples: CDT courses; CPD courses.] Other options. Please provide justification if using: Level 9/Year 3 [Deprecated except for compulsory UG3 courses. The course will not be available to other years.] Level 10/Year 4 Other:	
SCQF Credit Points	10x_20406080Other:	
Delivery Location	<u>x</u> CampusOn-line Distance Learning	
Course Type	<u>x</u> Standard (default)DissertationOnline Distance Learning PlacementStudent Led Individually Created CourseYear Abroac	
Marking Scheme	<u>x</u> Standard (numerical)Letter grade only Pass/Fail [Normally only for externally delivered courses]	

Guidance for remaining sections:

Before starting your proposal: please contact the DDoLT (Curriculum) informally before starting to complete this form, with at least the following information:

- Tentative course title, level, year, and number of credits
- Who the target audience is, and why the course is needed.

The DDoLT (Curriculum) or delegate will schedule a meeting with you to discuss your plans and whether a full course proposal makes sense. If so, you will be provided with further instructions.

Deadlines: New courses must be approved by the December BoS meeting to ensure allocation of teaching staff for the following academic year. Since it may require considerable discussion and iteration to prepare the proposal, you should **contact the DDoLT (Curriculum) as early as possible, ideally in spring or summer**, and you should **plan on submitting your full proposal by November.**

Submitting your proposal: When your proposal is complete, please submit to iss-bos@inf.ed.ac.uk.

Colour coding and item-by-item guidance:

Guidance is provided in italics for each item. Please also refer to the guidance for new course proposals at <u>http://www.inf.ed.ac.uk/student-services/committees/board-of-studies/course-proposal-guidelines.</u> Examples of previous course proposal submissions are available on the past meetings page <u>http://web.inf.ed.ac.uk/infweb/admin/committees/bos/meetings-directory</u> but note that the proposal form was updated in Apr 2021.

Sections in gold are for student view and are required before a course can be entered into DRPS.

Sections in orange are for School use but are still required for all courses (even those that have already been approved based on other documentation).

Section in gray are for consideration by the Board of Studies. They are normally required for all new course proposals but may be omitted in some cases, with permission (e.g., for invited proposals).

Glossary of terms:

(D)DoLT: (Deputy) Director of Learning and Teaching.

DRPS (<u>The Degree Regulations and Programmes of Study</u>): Provides the University's official listing and descriptions of courses, degree programmes, and the regulations that govern them; updated annually in April. Course information in DRPS is considered a contract with students.

DPT (Degree Programme Table): Lays out the course requirements for each year of a degree. All UoE degrees have a DPT in the DRPS.

<u>Path</u>: A system that students use to help choose courses and view options in their DPT. The information feeds through from DRPS but has a more student-friendly interface (e.g., by highlighting courses that are not running or where the student hasn't satisfied prerequisites).

SCQF (<u>The Scottish Credit Qualifications Framework</u>): Lays out the requirements for courses at different levels and with different numbers of credits.

1. Course overview and case for support

Except as noted, all fields are required and will go into the DRPS (course catalogue) entry for students. Important: Text in DRPS is effectively a contract with students, so should not include details that are likely to change from year to year.

Summary Description (for DRPS)

Since the early days of AI, researchers have been interested in making computers learn, rather than simply programming them to do tasks. This is the field of machine learning. The main area that will be discussed is supervised learning, which is concerned with learning to predict an output, given inputs. A second area of study is unsupervised learning, where we wish to discover the structure in a set of patterns, i.e. there is no output "teacher signal".

The aims of the course are to develop knowledge and a critical appreciation of the various techniques covered in the syllabus, and to be able to apply, validate and refine the methods in practice.

4

Contribution to curriculum; target audience and expected demand; consultation (for BoS only) Why is this course needed and how does it relate to existing courses and degree programmes (including any prerequisite courses)?

A decision has been taken to split the undergraduate (UG) and postgraduate (PG) courses that are currently both taught as IAML. This is necessary as the backgrounds of the UG and PG can be quite different. Our UG students will now have done Foundations of Data Science (FDS) in year 2, which includes some material on linear and logistic regression and statistical hypothesis testing, as well as aspects of data manipulation, summarization and visualization. Also our UG students have a stronger mathematical background than is required for the "math-lite" IAML course, which thus allows a more rigorous exposition. In addition, IAML has become very large (around 500 students in AY 2020/21) and splitting it will make the UG and PG courses more manageable.

What is the target audience, in terms of background and interests, and what is the expected demand (class size) for the course?

The level 10 (UG) enrolments in IAML for AY 2020/21 were around 250, with similar (although slightly lower) numbers in the preceding years.

Has this proposal been discussed with the DDoLT (Curriculum) or DoLT prior to BoS submission? ____ X Yes

No

Who else has been consulted?

A series of meetings were held with ML faculty to discuss the restructuring of IAML into PG and UG versions, and a general overview of ML courses.

Course Description (for DRPS)

This student-facing description should normally include (a) a more in-depth description of the learning aims, nature and context of the course, (b) a rough outline of the content, and (c) a description of how the course will be taught, and how students are expected to engage with it and to demonstrate their achievement of the learning outcomes. Note: Please keep this section general enough to avoid the need for yearly updates, and keep in mind that you should have only around 15 lecture hours of examinable material per 10pts of a course. (10pt courses may have 18-20 lecture hours, but the rest should be used for guest lectures, revision sessions, assignment feedforward/feedback, etc.)

The precise set of methods and algorithms used to illustrate and explore the main concepts will change slightly from year to year. However, the main topic headings are expected to be fairly stable.

- Classification and Regression: Linear Regression, logistic regression, Bayes classifiers
- Expanded feature representations: Basis functions, decision trees, neural networks, kernel methods
- Generalization and regularization: Penalised cost functions, Bayesian prediction, under- and over-fitting
- Model selection and performance evaluation: Cross-validation, ROC and PR curves
- Ethics and machine learning: Fairness, accountability, transparency, privacy concerns
- Representation learning: Dimensionality reduction, clustering, feature learning
- More advanced methods

To support these topics we will also cover:

- Optimization and Stochastic gradient descent
- Practical issues: Formulating problems as machine learning, adapting methods to fit problems. Numerical and programming issues important for machine learning.

Assessment Weightings (for DRPS)

These should correspond approximately to the proportion of learning outcomes (below) that each component assesses. Note that assessed coursework is typically more time-consuming than exams for both students and staff. A typical course is based no more than 30% on coursework and doing so requires justification.

Written Exam ____80___%

Practical Exam _____% (for courses with programming exams) Coursework ____20__%

Additional Information, Assessment (for DRPS)

State briefly for students what type of coursework to expect, including whether implementation is required. E.g., "Coursework will involve implementing some of the methods discussed" or "The coursework will assess students' analysis and proof skills. No implementation is required." More specific information can be useful, but please keep it high level and do not include details that are likely to change from year to year.

Coursework will involve use and possible modification of standard ML library functions, and critical evaluation of results obtained on provided datasets.

Learning Outcomes (MAXIMUM OF 5; for DRPS)

List the learning outcomes (LOs) of the course. These must be assessable (i.e., observable), so must specify what the student should be able to do concretely, not simply what they should "understand". Use concrete verbs that indicate (a) what type of assessment would be appropriate, and (b) what level of knowledge/thinking is expected (from recall to analysis to novel creation). **Example verbs:** define, explain, implement, compare, justify. Assessments (described later) should be tied to the LOs.

LOs should focus more on the types of thinking/skills developed than on the detailed course content, and should be appropriate to the level of the course: e.g., LOs at Level 11 should include more higher-level thinking skills than at Level 8. See <u>how to write good learning outcomes</u> and the <u>descriptors of the SCQF Levels</u>. Also, please consider including LOs related to **social or ethical implications** or **meta-skills** as well as technically-focussed LOs.

On completion of this course, the student will be able to

1) Explain the scope, goals and limits of machine learning (ML), and the main sub-areas of the field.

2) Describe and critically compare the various techniques covered in the syllabus, and explain where they fit within the structure of the discipline.

3) Demonstrate knowledge about the limitations and possible failure modes of ML models, and ethical issues in relation to machine learning.

4) Check and refine implementations of learning algorithms, and apply them in practice.

5) Use a systematic approach to conducting experimental investigations, including best practices on how to assess model performance.

Graduate Attributes, Personal & Professional Skills (for DRPS)

Please list the generic transferrable skills that this course will develop, as aligned with the <u>UoE's Graduate Attributes</u> framework. Examples from the four skills categories in the framework include:

- Research and enquiry: problem-solving, critical/analytical thinking, handling ambiguity, knowledge integration
- Personal effectiveness: leadership, planning and organizing, flexibility and change management, entrepreneurship
 Personal responsibility and autonomy: ethics and social responsibility, independent learning, self-awareness and
- reflection, creativity, decision-making
- Communication: interpersonal/teamwork skills; verbal, written, cross-cultural, or cross-disciplinary communication

The student will be able to reason about how to make predictions from and interpret data, an important transferable skill.

In addition the student will be able to:

Undertake critical evaluations of a wide range of numerical and graphical data.

Develop awareness of ethical issues in data analysis, and address associated ethical dilemmas.

Critically review and consolidate knowledge, skills, practices and thinking in subject/discipline/sector.

Commented [1]: Maybe add something about this, since it's a learning outcome?

2. Additional information on course design and resourcing (for BoS only, except where noted)

Breakdown of Learning and Teaching Activities (for DRPS) Please fill in the number of timetabled hours per student for each type of activity. Do not include non- timetabled hours.				
A typical 10pt Informatics course has:				
 18-20 lecture slots (2/wk), but only ~15h should be examinable lectures, with the rest used for guest lectures, revision sessions, assignment feedforward/ feedback, etc. If unsure of plans, count these under 'lecture hours' but please explain tentative plans in the free text below. 				
• No more than 4-5 lab or tutorial hours. Please consider whether fewer can be used, e.g. by using some lecture hours for whole-class discussion/feedforward.				
A typical 20pt course has 30 lecture slots (3/wk) and no more than 8 lab/tutorial hours. Timetabled Hours Type Lecture Hours: 30 Seminar/Tutorial Hours: 8 Dissertation Project Supervision Hours: 0 Supervised Lab/Workshop/Studio Hours: 0 Feedback/Feedforward hours: 2 Summative assessment hours: 2 [Normally 2h if using an exam; otherwise 0] Revision Session Hours: 1				
(Note for ISS: Remaining hours should be allocated to Directed and Undirected Learning Activities.)				
Use of timetabled activities (not to be included in DRPS)				

If labs or tutorials are planned, please describe their role in the course (e.g., as support for assessed coursework, review of exercises, discussion of ethical questions, etc). If a non-standard pattern or style of lectures is planned, please explain.

3 lectures per week for 10 weeks.

There will be 8 seminar/tutorial sessions (weeks 3-9), based on provided question sheets. We envisage a session size of around 48, divided up into groups of about 6 students, facilitated by 2 tutors. The groups will discuss the provided questions.

Summative assessment and time spent on assignments (not to be included in DRPS)

Please describe your plans for summative assessment, in more detail than in the student-facing description: How many and what types of assessment are planned (oral presentation, report, programming, etc)? For each piece of assessment, please indicate **(a)** which learning outcome(s) it assesses; and **(b)** how many hours students are expected to spend on it.

There will be one summative assignment, assessed by an individual report. This will involve completing a number of questions around data visualization and the use of ML methods on one or more datasets. It will include best practices on how to compare model performance, and potential modification of standard code.

This will assess Learning Outcome 4, and partly 5.

Hours: 20

Tentative plans for feedback/formative assessment (not to be included in DRPS)

Please describe your current plans for providing feedback to students: e.g. oral feedback during labs/tutorials, automarked solutions to in-lecture or online quizzes, peer feedback, etc. We also encourage submission of at least one piece of (individual or group) written work, with formative feedback emphasizing how students can improve.

- One formative assignment carried out in the first half of the course, similar in intent to the summative assignment, but limited by the content that will have been covered by then. Students will be permitted to work in small groups for this non-assessed coursework.
- For the assessed assignment, detailed feedback will be provided via the marking rubric (e.g. in Gradescope). We will also provide course level feedback during one of the lecture sessions.
- Oral feedback in seminar/tutorial sessions
- Students will have the opportunity to ask questions via an online forum (e.g. Piazza). They will be encouraged to provide peer feedback, but the instructors and TAs will also be available answer questions.

Commented [2]:

As I asked Oisin for the other course: does the exam also have some ability to assess these outcomes? (Perhaps 5, a bit?) If not, it seems reasonable to increase the cw weight a bit, unless the feeling is that these "practical" LOs are less important than the others.

Decolonisation and Inclusivity (not to be included in DRPS)

What actions are you taking towards making your course inclusive for all students, in terms of both **content** and **delivery**? Please be as specific as possible. If you are not taking any action, please justify. <u>See suggestions and guidance here</u>.

Content: The course includes material on Ethics in Machine Learning. This is in response to the fact that machine learning is now not just an academic subject, but is being used for problems that affect real people. We now see applications in healthcare, law enforcement, smart cities, retail, social media etc which may have direct effects on people's lives. The material covered builds on the general discussion of benefits and harms linked to data practices as covered in Foundations of Data Science (year 2), to cover fairness, accountability and transparency relating to machine learning. In particular fairness means that people are not discriminated against based on their membership of a protected group or class, such as race, gender, sexual orientation etc. The lecture materials include some examples of bias, e.g. in the data collection, and in assessing outcome fairness for differing values of a protected attribute.

Delivery: Using a purpose/task/criteria approach to coursework; further points to be addressed during course development.

Anticipated Resource Requirements If tutorials are needed, how many students per tutors? (Please provide your desired number, and the maximum feasible number.) 25 students/tutor (sessions of 50 handled by 2 tutors). This is much reduced relative to the standard 12 students per tutor, and cannot be further reduced. At 8 tutorials in the semester at an estimated capacity of 250 students, this is 8x 10 = 80 contact hours of tutor time If labs are needed, how many students per demonstrator? (Please provide your desired number, and the maximum feasible number.) No labs Please estimate the number of hours required for marking, per student. 1.25 hours per student (to mark the summative assignment), based on IAML experience. Exam marking: 0.5h per student If any other teaching support resource will be requested in order to develop or maintain the course, please provide an estimate of that here. TAs for Piazza and for assignment development (2 x 60h) Do you anticipate any difficulty recruiting enough teaching support? (For example if the course is very large or very specialized.) Possibly, given the size of the class. It was a struggle for IAML. Does the course have any scaling limits due to available space or equipment? No If equipment is required, please state how it will be procured and maintained. N/A Does the course have any external funding? (Typically only for CPD courses) No Does the course need any special arrangements such as quotas, agreements with other schools, or registration arrangements? Does it have any atypical characteristics that may affect finance or student registration? Please specify if so.

Not anticipated.

3. Further information for BoS consideration

A full proposal for a new course must include examples of exercises and assessment. Please provide these below, along with publicity information.

Course information and publicity

The course web page (typically the Learn landing page) will be linked from the Sortable Course List, and information such as timetables and assignment deadlines must be made available prior to the start of the academic year. Please specify here if any additional info/publicity is needed for your course, especially if it is aimed largely at non-SoI students.

Probably we should explain why the change was made from IAML, in order to better suit the UG cohort.

Sample tutorial/lab sheet questions

Provide a list of tutorial questions and answers and/or samples of lab sheets. These need not be fully fleshed out but should indicate what sort of exercises will be provided to help students learn the material.

See past IAML tutorial sheets, but note that the mathematical level for MLG will be slightly higher.

Sample assessment materials

If the course is primarily assessed by **exam**, provide a sample exam question with model answers. The <u>online list of past</u> <u>exam papers</u> gives an idea of typical and alternative exam formats.

If the course is largely or primarily assessed by **coursework**, provide a sketch of a possible assignment with an estimate of effort against each sub-task and a description of marking criteria.

See past IAML exam papers, but note that the mathematical level for MLG will be slightly higher. See e.g. the December 2018 paper

https://exampapers.ed.ac.uk/bitstream/handle/20.500.12593/103361/2019319_INFR11182.pdf?sequence=1&is Allowed=y]

Commented [4]: Same here – but also, since the course is mainly exam-based, we're looking for a sample exam question here. I can imagine that *some* questions might look like the examples Oisin included in the AML proposal, but as I pointed out there, those examples were missing a lot of the critical thinking aspects of the LOs. (Your LOs are slightly different, but the same comment applies.)

Commented [3]: But if the point is to differentiate from IAML, then presumably these will change somewhat?

4. Additional Course Details for DRPS

Except where otherwise noted, these fields are required for entry into EUCLID and will be visible to students in the DRPS entry.

Planned Academic Year of Delivery (The first year you anticipate the course running, e.g. AY 2019-20)	AY 2022-23	
Keywords Give a list of searchable keywords for the course.	Machine learning, supervised learning, unsupervised learning	
Course Organiser (By default, the course proposer)	TBD	
Intended Delivery Period	<u>x</u> Semester1 Semester 2 Full Year Summer Other (please specify):	
Timetable considerations/conflicts For School use. Please specify any constraints to be considered (e.g. overlap of popular combinations, other specialism courses, external courses etc). Include whether the semester delivery is constrained or could be flexible.		
Reading List/Learning Resources (for DRPS) You are encouraged to create resource lists using LEGANTO	 Books that may be useful, but are not required: Bayesian Reasoning and Machine Learning. David Barber (CUP, 2012) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Trevor Hastie, Robert Tibshirani and Jerome Friedman (Springer, 2009) Pattern Recognition and Machine Learning, Christopher Bishop (Springer, 2007) Machine Learning: A Probabilistic Perspective. Kevin P Murphy (MIT Press, 2012) 	
Feedback Information Provide a high-level description of how and what type of feedback will be provided to students, for inclusion in DRPS.	Oral feedback will be provided in tutorial sessions. Feedback at the course level will be provided for the assessed and non-assessed assignments. Piazza or a similar class discussion forum will be utilised for peer feedback.	

Is this course available to visiting students?	<u>x</u> Yes (default) No If no, please provide a justification here:		
Required pre-requisite courses Use sparingly: these are enforced in PATH and can only be waived by approval from the School's Curriculum Approval Officer. Note that cross-year required pre-requisites may prevent MSc students from registering; consider using recommended pre-requisites or "other requirements" instead.	 No Yes (please specify full course name(s) and code(s)): This course is intended for 3rd or 4th year Informatics students. Hence the following courses will have been taken Informatics 2 - Foundations of Data Science (INFR08030) and the maths courses Introduction to Linear Algebra (MATH08057), Calculus and its Applications (MATH08058), Discrete Mathematics and Probability (INFR08031) Similar maths courses can be substituted.] 	Commented [5]: If you say this, you may end up with loads of Maths or other CSE students (and they won't have done any of our intro courses, incl programming or algorithms). And IDS uses R.	
Recommended pre-requisite courses	<u>_x</u> No Yes (please specify full course name(s) and code(s)):	not Python. So is that really what you want? Given the size of the course, I would assume we want to limit it to Informatics degree students (incl joint degrees), at least for the first year.	
Required co-requisite courses Specify any courses that must be taken in parallel with the existing course. Note that this leads to a timetabling constraint that should be mentioned elsewhere in the proposal.	<u></u>		
Prohibited Combinations Specify any courses that may not be taken in combination with the proposed course].	 No x Yes (please specify full course name(s) and code(s)): Due to the high overlap of content, the courses Machine Learning for Pattern Recognition (INFR11130) and Introductory Applied Machine Learning (INFR11182) are a prohibited combination. 		

Other Requirements/Additional Information

This information is often used by MSc students and students from other Schools to see if they have appropriate background without having done our School's courses. So please avoid course titles, instead list specific knowledge and skills (such as mathematical concepts, programming ability or specific languages, etc).

Also list any other constraints on registration, for example: "Only available to 4th Year Informatics students including those on joint degrees." or "This course is open to all Informatics students including those on joint degrees, and to students in the School of Mathematics. Other external students whose DPT does not list this course should seek permission from the course organiser."

___No ___Yes (please specify):

This course is open to all Informatics undergraduate students including those on joint degrees. This course requires practical mathematical application of algebra, vectors and matrices, calculus, probability, and problem solving. Practical exercises usually require using a particular numerical language such as Python+NumPy. We will assume and require sufficient past programming experience that a new package can be learned on the fly. We also assume prior data science experience as per Foundations of Data Science (INFR08030) which covers data wrangling and exploratory data analysis, linear and logistic regression, statistical inference, and uses a Python-based ecosystem.

MSc students are not permitted to take this course, instead they should take Machine Learning for Pattern Recognition (INFR11130) or Applied Machine Learning (INFR11182).]

Visiting Student Pre-requisites

_x_Same as "other requirements" ____Different than "other requirements" (please specify):

Commented [6]:

See comment above. Maybe just not allow them?? (I'll need to check this with Bjoem.)

GOLDWATER Sharon This will be the new AML course instead, right?

5. Placement in degree programme tables: for level 9-11 courses only (except EPCC)

This section is for consideration by the Board of Studies and will be used later by ITO to determine where the course will be added to existing degree programme tables.

	Is this course restricted to students on a specific degree? E.g., some courses are only available to students on a specific CDT or MSc.	<u></u>
	Is this course compulsory for students on any degree(s)?	<u>x</u> No Yes (please specify and provide justification):
	Any issues for part-time students? Normally, part-time students have access to the same courses as full-time students on the equivalent degree. If you anticipate any problems with this, please specify here.	No

For optional courses:

If this course is available but non-compulsory for students on various degrees (most courses), please fill in this section. The choices here determine where the course appears in degree programme tables (DPTs) and the 2-3 character tags are displayed in the Informatics sortable course list.

Should this course be tagged as 'ML' (machine learning foundations and methods)? Courses with the ML tag are typically very high- demand and most degrees limit the number of ML credits. If your course might appeal to a similar audience but draw off students from these large courses, please select 'no' and choose one of the tags below.	No <u>x</u> Yes
If you chose 'no', please choose at least one of the following tags Ideally, select exactly one, unless there is a good argument for more than one. These three are used in various combinations for many of our degrees.	 FSS (CS foundations, systems, and software) AIA (artificial intelligence applications and paradigms) COG (cognitive science: including HCI and NLP courses, but not most other AI courses. Please restrict to courses most relevant to natural cognition.)

and also tick if any of the following tags or categories apply. Do not tick any of these if you selected 'ML' already.	 SE (software engineering: including courses that are highly relevant to SE degrees. All SE courses should also be FSS. This tag is mainly relevant for UG SE degrees.) Databases and data management systems (used for Data Science MSc and MSc(R)) Unstructured data and applications (used for Data Science MSc and MSc(R)) Level 11 Security courses (used for Security MSc) ATFC Optional courses (used for ATFC MSc)
If you are not sure which tags are most appropriate or have other questions about this section, please note any comments/issues here.	