The Hong Kong University of Science and Technology

Machine Learning for Ocean Science

OCES 5303

3 credits

Enrollment by instructor approval

Pre-requisites: Some programming (e.g. Python, R, Fortran, C, MATLAB…); having done the OCES Marine Genomics will help a bit with the Python; e-mail the instructor to discuss if in doubt

Name: Julian Mak

Email: jclmak@ust.hk

Consultation Hours: times TBC, Rm 5482 (Lift 25/26)

Course Description

Machine learning is increasingly prevalent in ocean sciences, and this course provides a hands-on introduction to some of the machine learning tools in the content of the related ocean science problems, in a progressive fashion in terms of complexity. Topics touched on include principles of regression and optimisation, and will focus quite significantly on the use of neural networks. The primary focus of the course will be on *utilising* the tools for ocean science applications, although some foray into the theory behind the tools will also be given.

The course will be "hands-on" in that the lectures are secondary to the computer workshops themselves: the delivery format is weekly sessions of three hours each in a computer barn, where the first 45-60 minutes will be an introduction to the concepts and tools, and the remaining time is for hands-on practice with the provided material that students work through at the own pace. The course material will be provided in Python; computing languages other than the one provided are not disallowed, but for practicality reasons, there will unlikely be official support from the course instructors on this (e.g., if you end up with syntax errors or bugs), and there is a strict demand on the assignments to be handed in the specified computing language. The course will be assessed via continuous assessment in the form of five quizzes for checking/assessing basic knowledge (every two weeks), two assignments and an extended project in the form of a scientific report for assessing practical competency, and interview style assessing mastery of subject.

Intended Learning Outcomes (ILOs)

By the end of this course, students should be able to:

1. Describe and explain how some problems in the ocean science context can be tackled through machine learning tools.

2. Perform calculations using machine learning tools to attack certain problems in marine science.

3. Evaluate and quantify the efficacy (e.g. skill) and limitations (e.g. interpretability) of the machine learning tools and its outputs for the relevant problem.

4. Illustrate the scientific case supported by the results from the machine learning calculations via a scientific presentations in the appropriate assessment format.

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided below, outlining the criteria used for evaluation.

Assessments:

The two written assessments are to be sent in as a Jupyter notebook (possibly including the relevant files) and submitted on Canvas. All Jupyter notebooks need to be able to run from scratch in a Conda environment at least, as this is part of the assessment on "reproducibility". We will maximally spend 10-15 mins on trying to fix bugs, but after that there will be penalisation on the coding and scientific content criteria (see below in grading rubric).

For the project, these should be in the form of a short paper as a pdf. The codes do not need to be provided (but you are encouraged to include it in line with open science policies). Your project will be assessed also during the interview style exam.

For all written assessments, students are allowed a 120 hour grace period with no questions asked (but you need to let us know on or before the assignment due time). Any further requests for extensions will require some proof (e.g. medical note, proof of internet going down such as electrical outage so Canvas submission was not possible).

For all types of assessments, there will be a late penalty of 1% per minute penalty (i.e. don't bother handing anything in after 100 mins, because you already got zero). Measure of lateness will be done via the Canvas timestamp.

Assessment Task	Contribution to Overall Course grade (%)	Due date
Assignment 1	15%	15th October 2025
Assignment 2	15%	12th November 2025
Project	25%	10th December 2025
Quizzes	20%	continuous
Interview style exam	25%	During exam period

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
Assignments 1, 2, project	ILO1, ILO2, ILO3, ILO4	All handed in reports require students to provide an outline and explanation of the context, perform their own calculations, and evaluate their results in a coherent manner fitting of a scientific report, and in line with expectations for capstone and FYPs. Additionally provides practice with report writing skills; appropriate feedback will be given.
Quizzes	ILO1, ILO2, ILO3	Quizzes provide continuous checks for students to evaluate their progress and understanding, and as a means to maintain engagement and promote class attendance.

Interview style exam	ILO1, ILO3, ILO4	The final oral exam serves as an overall assessment of the students' project as well as their mastery of the subject beyond technical competence in the execution aspect, in line with criterion-based grading.

Grading Rubrics

Quizzes are out of 10.

Written assessments are marked out of 110%, with the breakdown as scientific content (40%), writing and presentation (incl. grammar, figure quality etc.; 30%), coding and use of Jupyter notebooks functionalities (30%), and extras (going above and beyond the course material and approach; 10%). Most of the marks are given in terms of the scientific content to emphasise the focus, and the extras allows opportunity for further exploration. Any mark above 100% still only counts as 100%. Any false/broken references are penalised at 100% of the assignment (since this basically constituents academic misconduct).

Interview style exam is out of 25, with questions based on four randomly chosen topics drawn from a box by student on the day, and questions relating to the project will be asked. Questions depend on the combination of drawn topics.

For class attendance, 10 of the lectures are core and are the ones assessed with written assignments. Anything else are bonus lectures, which may be drawn for the interview style exam, but questions for those will be kept simple.

Final Grade Descriptors:

See also provided model good and bad hand-ins provided on the course GitHub page.

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance (>85% in the course)	Shows mastery of knowledge and understanding of the main subject matter, can problem-solve and critically evaluate approach, strong ability in communicating scientific and technical content.
B	Good Performance (60 – 85% in the course)	Shows good knowledge and understanding of the main subject matter, competence in problem-solving and some evaluation of approach, and the ability to communicate scientific and technical content.
C	Satisfactory Performance (50 – 60% in the course)	Shows adequate knowledge and understanding of the main subject matter, some issues with problem-solving, some ability to communicate scientific and technical content.
F	Fail (<50% in the course)	Shows insufficient knowledge and understanding of the main subject matter, struggles with problem-solving, unable to communicate scientific and technical content.

Course AI Policy

Use of AI is allowed and encouraged and by all means use it to help you code, but it is somewhat irrelevant in that most of the content is marked according to the scientific content anyway.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include a marked up PDF report with marked up comments, and a breakdown of the marks. Students who have further questions about the feedback including marks should consult the instructor within five working days after the feedback is received.

Resubmission Policy

Given the arrangements already for no-questions asked extensions and the use of continuous assessment, no resubmission or alternative assignments will be provided under normal circumstances.

Required Texts and Materials

All Jupyter notebooks should be self-contained; further reading is given in the notebooks accordingly.

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to Academic Integrity | HKUST – Academic Registry for the University's definition of plagiarism and ways to avoid cheating and plagiarism.

[Optional] Additional Resources

All Jupyter notebooks should be self-contained; further resources are given in the notebooks accordingly.

Proposed schedule

1. Introduction to machine learning (un-supervised vs. supervised, data types), refresher in coding language, introduction to scikit-learn, data processing and scaling, refresher in regression, probability and optimisation
2. Linear models, dimension reduction reduction, clustering, robustness, cross-validation, hyper-parameter tuning
3. Classification tasks, random forests, gradient boosting
4. Introduction to Neural Networks, introduction of PyTorch, building perceptrons (cf. ANNs)
5. Convolutional Neural Networks (CNNs)
6. Autoencoders
7. Recurrent Neural Networks (RNNs)
8. Generative Adversarial Networks (GANs)
9. Symbolic regression and equation discovery (SINDy)
10. Physics Informed Neural Networks (PINNs)

Three extra classes for preparation towards extended project presentation, with extra lectures on topics including but not limited to: GPU computing, usage of ML weather/climate forecasting models, Diffusion Models, Topological Data Analysis techniques in ML, issues of ethics, Fourier Neural Operators, control problems, Koopman operators, others as the field develops…