Module Template 2024-25

Module Code	Click here to enter text.
Module Name	**STATISTICAL SIGNAL PROCESSING AND MACHINE LEARNING (SPML)**
ECTS Weighting[1]	5 ECTS
Semester taught	Semester 1
Module Coordinator	Associate Professor ANTHONY QUINN
Module Learning Outcomes with reference to the Graduate Attributes and how they are developed in discipline	On successful completion of this module, students should be able to: **LO1** Analyze and design important signal processing and machine learning (SPML) algorithms using the principles of probability **LO2** Design tractable and efficient recursive computational flows for online filtering and prediction of standard engineering observation processes **LO3** Specify and rank alternative stochastic models for the purposes in LO1 and LO2 **LO4** Implement Wiener filters in standard stationary scenarios (filtering, equalization and system identification); and implement Kalman filters in nonstationary filtering scenarios **LO5** Compare parametric and nonparametric techniques for temporal and spatio-temporal regression problems **LO6** Derive optimal classifiers based on matched probability models, and compare them to off-the-shelf classifiers (*k*-means, EM) **LO7** Implement optimal transport (OT) solutions to problems of (i) resource allocation, and (ii) training-data repair for AI fairness (AIF) **Graduate Attributes: levels of attainment** To act responsibly - Attained To think independently - Attained To develop continuously - Enhanced To communicate effectively - Enhanced

Module Content	**OBJECTIVES** Statistical signal processing (SSP) has been a key enabler of the digital revolution for well over fifty years. It has provided vital algorithms for digital content generation and optimization, communication, and real-time decision-making. Moreover, the principles of SSP inform the designs of critical hardware infrastructures for data capture and processing. The foundations of SSP in linear algebra, signal theory, and statistical inference (particularly random processes) have yielded a powerful framework for supporting the design and analysis of ICT infrastructures. It is exciting now to witness the extent to which the same combination of design and analysis skills is impacting on the AI revolution, particularly in machine learning (ML). The optimal data-driven design of classifiers and ranking algorithms; the design of feature vectors for perceptual data sets; and the learning of regression structures via weights (e.g. embeddings) in a neural network: these are typical examples of ML tasks in which SSP principles can contribute usefully to their design and analysis. The aim of this module is to provide students with an understanding of key SSP principles that underlie ML tools encountered by any engineer working on data-driven problems today. Rather than catalogue and analyze those tools *per se*, the focus will be on their underlying SSP principles. As such, the module is aligned to the IEEE MLSP Technical Committee areas of interest, and aims to complement and support other machine learning and applied signal processing modules in the Engineering programme. The priority is to reveal to students the modelling assumptions and inference principles that allow SPML algorithms to work so well, and which also define their limits. For this reason, the module will review important principles in probability, statistics, and signal and system analysis, for the purposes of design, optimization and adaptation of important signal processing and machine learning (SPML) algorithms. A review of classical (least squares and likelihood-based) techniques, and their generalizations in Bayesian techniques, will provide students with an understanding of the design principles underlying Fourier-based signal analysis, hypothesis testing, linear regression, Bayesian classification, Wiener filtering, Kalman filtering and optimal transport. A priority is the design of recursive computational flows for efficient and real-time processing,

appreciating when such designs are consistent with the underlying model, and when not.

Two practical case studies will be addressed in the lectures and practical sessions: (i) Wiener and Kalman filter implementation for filtering and prediction of temporal and spatio-temporal engineering processes; and (ii) optimal transport (OT) methods for resource allocation at scale, and for repairing unfair training data in ML classifiers (**LO7**). In these contexts, students will be familiarized with available Matlab SPML tools, as well as with open-source Python tools available in TensorFlow and in the Python OT toolbox.

At the end of the module, students will understand the design principles underlying key SPML tools in the armory, their suitability for engineering applications, but also the basis for their adaptation and re-design, at the research frontier of this field.

SYLLABUS

- Review of the essentials of probability and random processes
- Bayesian and classical inferential techniques: posterior inference, and its specializations in likelihood, least squares and minimum mean squared error (MMSE) techniques
- Key parametric inference tasks: estimation, prediction and model inference (hypothesis testing)
- Linear (auto) regression, with applications in matched transform (Fourier and related) design and spectrum analysis
- Gaussian process regression, with applications to spatio-temporal processes
- Key online recursive algorithms: the Wiener and Kalman filters
- Classification of categorical data and clustered data (mixture modelling)
- Optimal transport: key progressions in data repair for AI fairness (**LO7**) and resource allocation

Teaching and Learning Methods

There will be approximately 44 contact hours. These will predominantly be delivered by the module coordinator, but there is an aspiration that two or three lectures will be provided by guest speakers. All sessions will be recorded and made available as a supplementary learning resource via Blackboard Collaborate Ultra (BbCU). A 3:1 ratio will be maintained between theory and tutorial content during the contact sessions. Problem-solving experience is vital. To this end, about six self-practice sheets will be distributed uniformly across the semester, and will form the basis for the tutorial content of the module.

A priority of the module is to demonstrate how SPML enables the design of tractable, real-time algorithms (**LO4** and **LO7**). Therefore, about 8 of the contact hours will be laboratory-based, involving the use predominantly of Matlab – but also Python – tools, with simulated data, and with AIF benchmark data sets (COMPAS, Adult, etc).

70% of the final module mark will be based on performance in the final examination. The remaining 30% will be reserved for continual assessment, by means of two in-semester tests, and a one-hour end-of-semester quiz.

Assessment Details

Assessment Component	Assessment Description	LOs Addressed	% of total	Week due
Two in-semester assessments	Written test (30-50 mins) Written / coding test (30-50 mins)	1-3 4,5	7.5 7.5	6 10
End-of-term quiz	Written test (1 hour)	1-7	15	12
Final examination	Written paper (2 hours)	1-7	70	Semester-1 examination week

Reassessment Requirements	There is no reassessment of this module.
Contact Hours and Indicative Student Workload	**Contact hours:** 44 **Independent Study (preparation for module and review of materials):** 44 **Independent Study (preparation for assessments, incl. completion of assessments):** 30
Recommended Reading List	The primary learning resource will be the full set of lecture notes developed in real time during the contact sessions, providing a unique record of the module. Audio-visual recordings of the sessions and their notes will be available for persistent access in BbCU. Supplementary learning materials, tutorial sheets, various self-study resources, etc., will also be distributed via Blackboard during the module. The following books may also support the student's learning in this module: Oppenheim, A.V. and Verghese, G.C., *Signals, Systems and Inference*, Global Edition, 1st ed., 2018. Proakis, J.G. and Manolakis, D.K., *Digital Signal Processing, 5*th ed., Pearson, 2021. Fessler, J.A. and Nadakuditi, R.R., *Linear Algebra for Data Science, Machine Learning and Signal Processing*. Cambridge University Press, 2024. Moon, T.K. and Stirling, W.C., *Mathematical Methods and Algorithms for Signal Processing.* Prentice Hall, 2000. McClellan, J.H., Burrus, C.S. *et al.,* *Computer-Based Exercises for Signal Processing.* Prentice Hall, 1998. Murphy, K.P., *Probabilistic Machine Learning: an Introduction.* MIT Press, 2022. Available online: https://probml.github.io/pml-book/book1.html

	Bishop, C.M., *Pattern Recognition and Machine Learning*. Springer, 2006. Available online: https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/ Rasmussen, C.E., *Gaussian Processes for Machine Learning.* MIT Press, 2006. Available online: https://gaussianprocess.org/gpml/chapters/RW.pdf
Module Prerequisite	The module is available to Engineering students who have performed well in mathematically-orientated modules. Indicative (but not required) modules include Digital Signal Processing (e.g. 4C5); Probability and Statistics (e.g. 3E3); Signals and Systems (e.g. 3C1); and Engineering Mathematics (up to Year 3 incl.). Interested students are invited to discuss the module with the coordinator, to assess whether they satisfy the prerequisites to enrol.
Module Co-requisite	None
Module Website	See module in Blackboard
Are other Schools/Departments involved in the delivery of this module? If yes, please provide details.	No
Module Approval Date	
Approved by	
Academic Start Year	
Academic Year of Date	2024-25