H7AML: Advanced Machine Learning

Module Code:	H7AML
Long Title	Advanced Machine Learning APPROVED
Title	Advanced Machine Learning
Module Level:	LEVEL 7
EQF Level:	6
EHEA Level:	First Cycle
Credits:	10
Module Coordinator:	Arghir Moldovan
Module Author:	Arghir Moldovan
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	MSc and/or PhD degree in computer science or cognate discipline. May have industry experience also.

Learning Outcomes

On successful completion of this module the learner will be able to:

#	Learning Outcome Description
LO1	Apply and evaluate the efficacy of advanced data preparation methods
LO2	Build and apply advanced methods for prediction and forecasting in various problem domains
LO3	Build and evaluate advanced machine learning models in various problem domains
LO4	Extract, interpret and analyse information and knowledge from non-trivial real-world datasets
LO5	Summarise, critique and present the results of advanced machine learning models in various problem domains

Dependencies

Module Recommendations

67243	H6DMML	Data Mining and Machine Learning

Co-requisite Modules

No Co-requisite modules listed

Entry requirements	learners should have attained the knowledge, skills and competence gained from stage 2 of the BSc (Hons) in Data Science

H7AML: Advanced Machine Learning

Module Content & Assessment

Indicative Content
General Strategies Revisited Increasing data complexity and size with fundamental methods. Considerations of Complexity on Computing Requirements
General Strategies Revisited Dimensionality Reduction (PCA, MCA, etc.). Feature Engineering. Measuring Predictor Importance
General Strategies Revisited Understanding, Detecting and Handling (massive) class imbalance. Understanding Factors that can Affect Model Performance; e.g. Type III errors, selection bias, measurement errors, improper variable encoding. Ethically assessing biases.
Advanced Regression Models Regression revision, and penalised models
Advanced Regression Models Generalised Linear Modelling. Automated Linear Modelling via Bagging and Boosting
Ensembles Ensembles:. Random Forest. Voting. Stacking. Bagging and Boosting Methods (e.g. XGBoost, AdaBoost, CART aggregation etc.)
Black Box Methods Support Vector Machines and Support Vector Regression
Black Box Methods Neural Networks: Classic Topologies and Activation Functions; Back Propagation; Gradient Descent and Stochastic Gradient Descent; Hyperparameter Optimisation techniques
Deep Regression Models A brief introduction to deep learning applied to regression problems (e.g. GLMNet). Special emphasis to be played on when these methods are(n't) appropriate (e.g. data volumes required).

Assessment Breakdown	%
Coursework	100.00%

Assessments

Full Time

Coursework

Assessment Type:	Continuous Assessment	**% of total:**	Non-Marked
Assessment Date:	n/a	**Outcome addressed:**	2,3
Non-Marked:	Yes		

Assessment Description:
Ongoing feedback on ongoing lab activities.

Assessment Type:	Project	**% of total:**	50
Assessment Date:	n/a	**Outcome addressed:**	3,4
Non-Marked:	No		

Assessment Description:
Team project; applying methods of the module to real world datasets such as Kaggle, Dublinked.ie etc.

Assessment Type:	Easter Examination	**% of total:**	50
Assessment Date:	n/a	**Outcome addressed:**	1,2,5
Non-Marked:	No		

Assessment Description:
Individual Hackathon where learners identify the fit and/or appropriateness of a variety of methods to one or more appropriately sized datasets.

No End of Module Assessment

No Workplace Assessment

Reassessment Requirement

Repeat examination
Reassessment of this module will consist of a repeat examination. It is possible that there will also be a requirement to be reassessed in a coursework element.

Reassessment Description
The repeat strategy for this module is a terminal assessment. Students will be afforded an opportunity to repeat the assessment at specified times throughout the year and all learning outcomes will be assessed in the repeat assessment.

H7AML: Advanced Machine Learning

Module Workload

Module Target Workload Hours 0 Hours

Workload: Full Time

Workload Type	*Workload Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Classroom & Demonstrations (hours)	24	Per Semester	2.00
Tutorial	Other hours (Practical/Tutorial)	24	Per Semester	2.00
Independent Learning	Independent learning (hours)	202	Per Semester	16.83
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

Hastie, T., Tibshirani, R. & Friedman, J.. (2016), The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed), Springer Series in Statistics.

James, G., Witten, D., Hastie, T. & Tibshirani, R.. (2017), An Introduction to Statistical Learning: with Applications in R, Springer Texts in Statistics.

Kuhn, M. & Johnson, K.. (2013), Applied Predictive Modeling, Springer.

Shalev-Shwartz, S. & Ben-David, S.. (2014), Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press.

Supplementary Book Resources

Downey, B.. (2014), Think Stats: Exploratory Data Analysis, (2nd ed).

Goodfellow, I., Bengio, Y., & Courville, A.. (2016), Deep Learning, The MIT Press.

Hearty, J.. (2016), Advanced Machine Learning with Python, Packt Publishing Ltd.

Leskovec, J. Rajaraman, A., & Ullman, J.. (2014), Mining of Massive Datasets, Cambridge University Press.

Wickham, H. & Grolemund, G.. (2017), R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, O'Reilly.

This module does not have any article/paper resources

Other Resources

http://www.datacamp.com

http://www.kdnuggets.com

http://www.r-bloggers.com

Discussion Note:	