

Lingnan University
Course Syllabus

Course Title	:	Machine Learning
Course Code	:	CDS3003
Recommended Study Year	:	3
No. of Credits/Term	:	3
Mode of Tuition	:	Sectional Approach
Class Contact Hours	:	3 hours per week
Category in Major Prog.	:	Required
Prerequisite(s)	:	(a) CDS2001 Probability and Statistics 2, (b) CDS2002 Introduction to Artificial Intelligence, and (c) SSC2113 Linear Algebra
Co-requisite	:	Nil
Exclusion	:	Nil
Exemption Requirement	:	Nil

Brief Course Description:

This course is designed to introduce methods in machine learning (i.e., the study of how to create a machine that is capable of improving its performance in some tasks, such as perception, cognition and action, by learning from experience and data). This makes computers more intelligent. Machine learning is a powerful tool for recognising hidden and potentially useful relationships in data. It is essential in many organisations, as they can generate huge amounts of data in their daily operations and machine learning enables them to discover important knowledge from this data. The students will learn some powerful software packages for performing machine learning including generative artificial intelligence (AI) models and apply these packages to handle different kinds of problems, culminating in a group project to solve a real-world problem.

Aims:

This course aims at delivering the concepts and applications of machine learning. The course discusses different categories of machine learning algorithms. The course will cover the supervised learning, unsupervised learning, reinforcement learning, and neural networks. The basic concepts and applications of reinforcement learning will be discussed. The course also aims at introducing some Python packages for machine learning, such as scikit-learn and Pytorch, and enables students to apply different learning algorithms available in these packages. Moreover, the theories and techniques for data processing and model selection will be introduced throughout the course.

Learning Outcomes (LOs):

Upon the successful completion of this course, the student will be able to:

1. Identify different categories of machine learning algorithms;
2. Illustrate and explain supervised learning, unsupervised learning, reinforcement learning and neural networks (PLO5);
3. Describe and apply the theories and techniques for data processing and model selection;
4. Develop machine learning programs to solve some real-world problems (PLO9);
5. Develop project management and teamwork skills.

Indicative Contents:

Machine Learning

Basic concepts of machine learning

Different categories of learning algorithms

Supervised Learning

The concepts of supervised learning

Regression, classification

Unsupervised Learning

The concepts of unsupervised learning including generative artificial intelligence
generative adversarial networks, diffusion models, language models

Reinforcement Learning

The basic concepts of reinforcement learning

Dataset Manipulation and Preprocess

Dataset manipulation, dataset preprocess

Model Selection

Confusion matrix, F-measure, regularization, bias-variance trade-off

Neural Networks

The concepts and applications of neural networks

Linear neural networks, multilayer perceptrons, convolutional neural networks

Project Management

How to work in a team on a datathon project

Use of Github and Google Colab for coding collaborations

Teaching Method:

There are different teaching and learning activities including lectures, tutorials, and project-based learning. The concepts of machine learning and different categories will be discussed in lectures. Students will also learn various supervised learning and unsupervised learning methods in lectures. The implementations of some learning algorithms in Python will be discussed in tutorials. Moreover, students will also understand the implementations of different algorithms in some machine learning software packages such as scikit-learn and Pytorch. The detailed elaborations and computations of some learning algorithms will be discussed in tutorials. Finally, students will work on a group project.

Measurement of Learning Outcomes:

	Class Attendance and Participation	Assignments	Mid-term Examination	Group Project
1. Identify different categories of machine learning algorithms			x	x
2. Illustrate and explain supervised learning, unsupervised learning, reinforcement learning and neural networks	x	x	x	x
3. Describe and apply the theories and techniques for data processing and model selection			x	x
4. Develop machine learning programs to solve some real-world problems		x		x
5. Develop project management and teamwork skills				x

1. There are a number of classroom activities to evaluate if the students can describe supervised learning, unsupervised learning, reinforcement learning, and model selection methods (LO2 and PLO5).
2. Written assignments require students to derive some parts of mathematical theory about the machine learning methods and/or perform related calculations (LO2 and PLO5).
3. Programming assignments require students to implement a prototype of machine learning system to solve some problems using computers. Students will acquire the IT skills related to machine learning (LO4 and PLO9).
4. Understanding of different types of machine learning methods and model selection techniques is comprehensively assessed in the examinations (LO1-3 and PLO5).
5. Group projects require that students collaborate effectively to apply machine learning techniques to solve real-world problems (LO5 and PLO3-6).

Assessment:

Class Attendance and Participation	15%
Assignments	40%
Mid-term Examination	15%
Group Project	30%
Total	100%

Required/Essential Readings:

1. Domingos, Pedro. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books, 2015.
2. Shalev-Shwartz, S. and Ben-David, S. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge university press, 2014.
3. Foster, David. *Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play*. O'Reilly Media, 2019.

Recommended/Supplementary Readings:

1. Bishop, Christopher M., *Pattern Recognition and Machine Learning*, Springer, 2008.
2. Géron, Aurélien. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc., 2022.
3. Paleyes A, Urma R G, Lawrence N D. Challenges in deploying machine learning: a survey of case studies[J]. *ACM computing surveys*, 2022, 55(6): 1-29.
4. Kelleher, John D., Brian Mac Namee, and Aoife D'arcy. *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT press, 2020.
5. Mitchell, Tom M., *Machine Learning*, McGraw-Hill, 1997.
6. Murphy, Kevin P., *Machine Learning: A Probabilistic Perspective*, MIT press, 2012.
7. Raschka, Sebastian, and Vahid Mirjalili. *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*. Packt publishing ltd, 2019.
8. Sutton, Richard S. and Barto, Andrew G., *Reinforcement Learning: An Introduction*, MIT Press, 1998.
9. Brown, Tom B., et al. "Language Models are Few-Shot Learners." In Proceedings of the 2020 Conference on Neural Information Processing Systems (NeurIPS 2020).
10. Liang P P, Zadeh A, Morency L P. Foundations & trends in multimodal machine learning: Principles, challenges, and open questions[J]. *ACM Computing Surveys*, 2024, 56(10): 1-42.

Important Notes:

- (1) Students are expected to spend a total of 9 hours (i.e. 3 hours of class contact and 6 hours of personal study) per week to achieve the course learning outcomes.
- (2) Students shall be aware of the University regulations about dishonest practice in course work, tests and examinations, and the possible consequences as stipulated in the Regulations Governing University Examinations and Course Work. In particular, plagiarism, being a kind of dishonest practice, is “the presentation of another person’s work without proper acknowledgement of the source, including exact phrases, or summarised ideas, or even footnotes/citations, whether protected by copyright or not, as the student’s own work”. Students are required to strictly follow university regulations governing academic integrity and honesty.
- (3) Students are required to submit writing assignment(s) using Turnitin.
- (4) To enhance students’ understanding of plagiarism, a mini-course “Online Tutorial on Plagiarism Awareness” is available on <https://pla.ln.edu.hk/>.
- (5) Students must adhere to the University’s guidelines and practices when using Generative Artificial Intelligence (AI) tools. The official documents “Guidelines for Using GAI Tools at Lingnan University” and “Best Practices for Ethical and Responsible Use of GAI Tools in Course Assessments” can be found here: <https://www.ln.edu.hk/tlc/generative-artificial-intelligence/gai-guidelines-best-practices>.

Rubric for Mid-Term Examination of CDS3003 - Machine Learning

Criteria	Very good (4-6)	Satisfactory (2-4)	Unsatisfactory (0-2)
Identify different categories of machine learning algorithms (LO1)	Demonstrates a strong capability of identifying appropriate machine learning algorithms.	Demonstrates an adequate level of capability of identifying appropriate machine learning algorithms.	Fails to identify appropriate machine learning algorithms.
Illustrate supervised learning, unsupervised learning, and neural networks (LO2)	Demonstrates a strong capability of illustrating appropriate machine learning algorithms.	Demonstrates an adequate level of capability of illustrating appropriate machine learning algorithms.	Fails to illustrate machine learning algorithms.
Apply the concept related to model selection (LO3)	The student can create different confusion matrix for some supervised learning problems. The student can calculate the values of all measures. More than 90% of these values are correct.	The student can create different confusion matrix for some supervised learning problems. The student can calculate the values of all measures. Between 60%- 89.99% of these values are correct.	The student cannot create different confusion matrix for some supervised learning problems. The student cannot calculate the values of many measures. Less than 60% of these values are correct.
Presentation	Content of submission/ presentation is well formatted with a clearly readable layout and no/very few grammatical/ formatting mistakes.	Content of submission/ presentation is properly formatted with a reasonable layout and no more than a few grammatical/ formatting mistakes.	Content of submission/ presentation is not properly formatted and/or there are more than a few grammatical/ formatting mistakes.

Rubric for Written Assignments of CDS3003 - Machine Learning

Criteria	Very good (4-6)	Satisfactory (2-4)	Unsatisfactory (0-2)
Perform calculations that are related to the theory of Decision Tree. (LO2)	More than 80% of the calculations are correct.	Between 50% and 80% of the calculations are correct.	Less than 50% of the calculations are correct.
Perform calculations that are related to Logistic Regression. (LO2)	More than 80% of the calculations are correct.	Between 50% and 80% of the calculations are correct.	Less than 50% of the calculations are correct.
Derive some parts of mathematical theory about some unsupervised learning methods. (LO2)	The student correctly derives these parts of mathematical theory.	The student derives these parts of mathematical theory with some errors.	The student cannot derive these parts of mathematical theory.

Rubric for Programming Assignments of CDS3003 - Machine Learning

Criteria	Very good (4-6)	Satisfactory (2-4)	Unsatisfactory (0-2)
Problem definition (demonstrate the understanding of the problem and consider alternative solutions) (LO4)	Clearly states the problem, lists related constraints, and identifies alternative solutions.	The problem is stated but related constraints and alternative solutions are not considered thoroughly.	The problem is marginally defined and with no consideration of constraints and alternative solutions.
Design of the machine learning system (LO4)	The design of the system is strongly related to the problem, The student can clearly explain the approaches and techniques involved. Moreover, the design of the system is very innovative.	The design of the system is related to the problem. The student does not clearly explain the approach and techniques involved. Moreover, the design of the system is not very innovative.	The design of the system is weakly related to the problem. The student does not clearly explain the approach and techniques involved. Moreover, the design of the system is not innovative.
Implementation of the machine learning system (LO4)	The machine learning system can be implemented correctly. The efficiency and effectiveness of the system are excellent.	The machine learning system is implemented with some minor errors. The efficiency and effectiveness of the system are acceptable.	The machine learning system cannot be implemented or the implementation(s) has too many errors. The system cannot be executed, or the efficiency and effectiveness of the system are poor.
Code Quality	Produces efficient, clean, and well-documented code that demonstrates a superior understanding of programming best practices.	Code is generally understandable and organized.	Code is disorganised and hard to follow.

Rubric for Group Project of CDS3003 - Machine Learning

Criteria	Very good (4-6)	Satisfactory (2-4)	Unsatisfactory (0-2)
Identify different categories of machine learning algorithms (LO1)	Demonstrates a strong capability of identifying appropriate machine learning algorithms.	Demonstrates an adequate level of capability of identifying appropriate machine learning algorithms.	Fails to identify appropriate machine learning algorithms.
Illustrate supervised learning, unsupervised learning, reinforcement learning, and neural networks (LO2)	Demonstrates a strong capability of illustrating appropriate machine learning algorithms.	Demonstrates an adequate level of capability of illustrating appropriate machine learning algorithms.	Fails to illustrate machine learning algorithms.
Describe and apply the theories and techniques for model selection (LO3)	Demonstrates a strong capability to know which models are appropriate for different datasets and problems.	Demonstrates an adequate level of capability to know which models are appropriate for different datasets and problems.	Fails to select the correct model for different machine learning problems.
Develop machine learning programs to solve some real-world problems (LO4)	Demonstrates a strong capability to write and test a computer program that accurately solves a real-world machine learning problem.	Demonstrates an adequate level of capability to write and test a computer program that accurately solves a real-world machine learning problem.	Demonstrates a lack of understanding of how to write a computer program to solve a real-world machine learning problem.
Presentation	Content of submission/presentation is well formatted with a clearly readable layout and no/very few grammatical/formatting mistakes.	Content of submission/presentation is properly formatted with a reasonable layout and no more than a few grammatical/formatting mistakes.	Content of submission/presentation is not properly formatted and/or there are more than a few grammatical/formatting mistakes.

Rubric for Class Attendance and Participation of CDS3003 Machine Learning

Criteria	Very good (4-6)	Satisfactory (2-4)	Unsatisfactory (0-2)
Attendance	Full, punctual attendance in class and mandatory seminars.	Occasional absences or lateness from class or mandatory seminars.	Frequent or recurring absence or lateness from class or mandatory seminars.
Class Participation	Active class participation and leadership in group activities. Demonstrates deep understanding of course content.	Passive class participation and active in group activities. Demonstrates adequate understanding of course content.	Lack of participation and active disruption of class and group activities.