Semester 8
8.1 Reinforcement Learning

School of Business, IT and Management

Reinforcement Learning

Insert Academic Year

School-Program	Year	Semester
Honours Bachelor of Artificial Intelligence	4	8

Course Code: Click or tap here to enter text.	**Course Equiv. Code(s)**: Click or tap here to enter text.	
Total Course Hours: 42 **Lecture Hours**: 3 **Lab/Tutorial Hours**: N/A	**Credit Value**: 3	
Prerequisite: Introduction to Machine Learning, Introduction to Artificial Neural Networks		
Corequisite: N/A		
Laptop Course: **Yes** ☒	**No** ☐	
Delivery Mode(s): **In class** ☒	**Online** ☐	**Hybrid** ☐
Authorized by (Dean):	**Date:**	

Prepared by: Dr. Ehsan Amjadian, PhD in Cognitive Science, Natural Language Processing, Deep Learning
Qualified to teach: PhD in Cognitive Science, Natural Language Processing, Deep Learning, Machine Learning, Computer Science

Course Description:

This course is an introduction to the field of reinforcement learning (RL). RL is primarily concerned with understanding how a software agent learns to behave in an environment to maximize the reward. Students explore methods, algorithms, and theoretical and practical aspects of RL, including exploration and generalization, the definition of state space, action space, dynamics, rewards, on-policy and off-policy learning, value iteration and policy iteration for RL.

Course Learning Outcomes:

Course Specific Learning Outcomes (CLO) Students receiving a credit for this course will have demonstrated their ability to:
1. Explain the functions of the terms and elements of reinforcement learning formalisms, notations, and equations.
2. Explain the circumstances when reinforcement learning is applicable.
3. Discuss the advantages and disadvantages of reinforcement learning by contrasting to other areas in machine learning.
4. Code reinforcement learning algorithms to automatically achieve a goal having learnt from rewards and the absence of rewards.
5. Analyze the architecture of deep reinforcement learning models to determine modifications required to improve performance on different tasks.
6. Analyze the characteristics of various reinforcement learning paradigms to determine the appropriate application.

Students will be notified in writing of changes that involve the addition or deletion of learning outcomes or evaluations, prior to changes being implemented, as specified in the Course Outline Policy and Procedure at Durham College.

Undergraduate Degree Standards:

This course will contribute to the achievement of the following Undergraduate Degree Standards, as outlined in the Ontario Qualifications Framework.

Undergraduate Degree Standards	**Course Learning Outcome Reference**
Depth and Breadth of Knowledge	CLO6
Conceptual & Methodological Awareness/Research and Scholarship	CLO3

Communication Skills	CLO1
Application of Knowledge	CLO4
Professional Capacity/Autonomy	CLO5
Awareness of Limits of Knowledge	CLO2

Program Learning Outcomes:

This course will contribute to the achievement of the following Program Learning Outcomes:

1.	Evaluate data requirements and technical approaches for building an AI solution by using systems analysis to align to business and client concerns.
3.	Build machine learning models by evaluating input data and identifying features that meet the needs of the project.

Evaluation Criteria:

Evaluation	Course Learning Outcomes	Weighting
Quiz 1	CLO1, CLO2, CLO6	15%
Assignment 1	CLO4, CLO5, CLO6	10%
Quiz 2	CLO1, CLO3, CLO6	15%
Assignment 2	CLO4, CLO5, CLO6	10%
Quiz 3	CLO1, CLO6	15%
Quiz 4	CLO1, CLO5, CLO6	15%
Term Project	CLO1, CLO2, CLO3, CLO4, CLO5, CLO6	20%
Total		**100%**

Students will be notified in writing of changes that involve the addition or deletion of learning outcomes or evaluations, prior to changes being implemented, as specified in the Course Outline Policy and Procedure at Durham College.

Required Text(s) and Supplies:

1. RT1: Sutton, R. S., Barto, A. G. (2018). Reinforcement Learning: An Introduction. The MIT Press.
2. RT2: Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M.A. (2013). Playing Atari with Deep Reinforcement Learning. ArXiv, abs/1312.5602.
3. RT3: Hasselt, H.V., Guez, A., & Silver, D. (2015). Deep Reinforcement Learning with Double Q-learning. AAAI.
4. RT4: Adam Paszke Reinforcement Learning DQN (PyTorch) Tutorial.

Classroom Equipment and Requirements:

1. Students need access to cloud computing (Amazon AWS, Microsoft Azure, or Google Cloud Platform) in order to implement these techniques most efficiently.
2. Computers to access the cloud resources mentioned above.

Recommended Resources (purchase is optional):

1. Lapan, M. (2018). Deep Reinforcement Learning Hands-On. Birmingham, UK: Packt Publishing. ISBN: 978-1-78883-424-7
2. RT4: Fujimoto, S., Hoof, H.V., & Meger, D. (2018). Addressing Function Approximation Error in Actor-Critic Methods. ICML.

Learning Plan	
Week	**Lecture** **Hours:**
1	**Topics:** ● Course Syllabus & Outline ○ Course Materials ○ Topics per session ○ Evaluation ● What is Reinforcement Learning ○ Components of Reinforcement Learning ○ Examples ○ Early History ● Multi-armed Bandits Part 1 **Intended Learning Objectives:** ● Understand the RL problem ● Analyze the Multi-Armed Bandit problem
	Intended Learning Activities: ● Lecture
	Resources and References: ● RT1 Chapter 1 and 2
	Evaluation and Weighting: N/A
	Course Learning Outcome Reference: CLO3
Week	**Lecture** **Hours:**
2	**Topics:** ● Multi-armed Bandits Part 2 ● Finite Markov Decision Processes Part 1 **Intended Learning Objectives:** ● Evaluate various multi-armed bandit problems and their corresponding solutions ● Understanding the components of a Markov Decision Process

	Intended Learning Activities: ● Lecture, in-class exercise
	Resources and References: ● RT1 Chapter 2 and 3
	Evaluation and Weighting: N/A
	Course Learning Outcome Reference: CLO1, CLO6
Week	**Lecture** **Hours:**
3	**Topics:** ● Finite Markov Decision Processes Part 2 ● Dynamic Programming **Intended Learning Objectives:** ● Analyze Markov Decision Processes ● Evaluate Dynamic Programming Algorithms in computing optimal policies given a perfect model of the environment as a Markov decision process
	Intended Learning Activities: ● Lecture, in-class exercise
	Resources and References: ● RT1 Chapter 3 and 4
	Evaluation and Weighting: N/A
	Course Learning Outcome Reference: CLO2
Week	**Lecture** **Hours:**
4	**Topics:** ● Quiz 1 ● Monte Carlo Methods **Intended Learning Objectives:**

	● Evaluate Monte Carlo Methods in finding optimal policies at the absence of complete knowledge of the environment
	Intended Learning Activities: ● Lecture, Quiz
	Resources and References: ● RT1 Chapter 5
	Evaluation and Weighting: Quiz 1 (15%)
	Course Learning Outcome Reference: CLO1, CLO3
Week	**Lecture** **Hours:**
5	**Topics:** ● Temporal Difference Learning **Intended Learning Objectives:** ● Evaluate TD Learning regarding how it approaches the prediction problem ● Compare and contrast TD against Monte Carlo Methods and Dynamic Programming
	Intended Learning Activities: ● Lecture, in-class exercise
	Resources and References: ● RT1 Chapter 6
	Evaluation and Weighting: N/A
	Course Learning Outcome Reference: CLO6
Week	**Lecture** **Hours:**
6	**Topics:** ● n-step Bootstrapping **Intended Learning Objectives:**

	• Contrast one-step TD with n-step TD generalizing TD as well as Monte Carlso methods.
	Intended Learning Activities: • Lecture, in-class exercise
	Resources and References: • RT1 Chapter 6
	Evaluation and Weighting: Assignment 1 (10%)
	Course Learning Outcome Reference: CLO1, CLO2, CLO3, CLO4
Week	**Lecture** **Hours:**
7	**Topics:** • Model-based and Model-free Reinforcement Learning **Intended Learning Objectives:** • Compare and contrast planning and learning in reinforcement learning • Evaluate model-based and model-free approaches in computing a value function
	Intended Learning Activities: • Lecture, in-class programming exercise
	Resources and References: • RT1 Chapter 8
	Evaluation and Weighting: N/A
	Course Learning Outcome Reference: CLO1, CLO6
Week	**Lecture** **Hours:**
8	**Topics:** • Quiz 2 • Approximating On-policy Prediction **Intended Learning Objectives:**

	● Analyze function approximation in reinforcement learning to estimate a state-value function. ● Contrast function approximation methods in RL with tabular reinforcement learning	
	Intended Learning Activities: ● Lecture, in-class programming exercise	
	Resources and References: ● RT1 Chapter 9	
	Evaluation and Weighting: Quiz 2 (15%)	
	Course Learning Outcome Reference: CLO3, CLO6	
Week	**Lecture**	**Hours:**
9	**Topics:** ● On-policy Control with Approximation **Intended Learning Objectives:** ● Evaluating the control problem with a parametric approximation of the action-value function in on-policy RL	
	Intended Learning Activities: ● Lecture	
	Resources and References: ● RT1 Chapter 10	
	Evaluation and Weighting: N/A	
	Course Learning Outcome Reference: CLO1, CLO6	
Week	**Lecture**	**Hours:**
10	**Topics:** ● Off-policy Methods with approximation **Intended Learning Objectives:**	

	● Contrast off-policy with function approximation with its on policy counterparts ● Analyze off-policy RL with approximation in terms of convergence and performance.
	Intended Learning Activities: ● Lecture, in-class programming
	Resources and References: ● RT1 Chapter 11
	Evaluation and Weighting: Assignment 2 (10%)
	Course Learning Outcome Reference: CLO3, CLO4, CLO6
Week	**Lecture** **Hours:**
11	**Topics:** ● Eligibility Traces **Intended Learning Objectives:** ● Lecture, in-class exercise
	Intended Learning Activities: ● Analyze the spectrum created by the eligibility traces in order to recognize the extreme poles and find the optimal methods ● Evaluate how eligibility traces combine Monte Carlo methods and TD methods
	Resources and References: ● RT1 Chapter 12
	Evaluation and Weighting: N/A
	Course Learning Outcome Reference: CLO6
Week	**Lecture** **Hours:**
12	**Topics:** ● Quiz 3

	● Policy Gradient Methods **Intended Learning Objectives:** ● Evaluate parametrized policies that directly select actions without consulting a value function. ● Analyze how might value functions interact with such parameterized policies.
	Intended Learning Activities: ● Lecture
	Resources and References: ● RT1 Chapter 13
	Evaluation and Weighting: Quiz 3 (15%)
	Course Learning Outcome Reference: CLO4, CLO6
Week	**Lecture** **Hours:**
13	**Topics:** ● Neural Actor-Critic ● TD-Gammon ● DQN ● Double Q-learning **Intended Learning Objectives:**
	Intended Learning Activities: ● Analyze TD-Gammon and modern neural Q-learning, comparing the two paradigms, advantages and disadvantages
	Resources and References: ● RT1 Chapter 16 ● RT2 ● RT3
	Evaluation and Weighting: N/A

	Course Learning Outcome Reference: CLO5, CLO6	
Week	**Lecture**	**Hours:**
14	**Topics:** ● Term Project Due ● DQN with Pytorch ● Quiz 4 **Intended Learning Objectives:** ● Applying the PyTorch framework to implement deep reinforcement learning algorithms.	
	Intended Learning Activities: ● Facilitated group coding workshop ● Question & Answer	
	Resources and References: ● RT4	
	Evaluation and Weighting: Quiz 4 (15%) Term Project (20%)	
	Course Learning Outcome Reference: CLO4, CLO6	