

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра штучного інтелекту та програмного забезпечення

“ЗАТВЕРДЖУЮ”

Проректор
з науково-педагогічної роботи



А.В. Пантелеймонов

2018 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Розробка систем штучного інтелекту

рівень вищої освіти	другий (магістерський) рівень
галузь знань	12 Інформаційні технології
спеціальність	122 Комп’ютерні науки
освітня програма	Інформаційні управляючі системи і технології
спеціалізація	
вид дисципліни	обов’язкова
факультет	комп’ютерних наук

Програму рекомендовано до затвердження Вченою радою факультету комп'ютерних наук
«05» вересня 2018 року, протокол № 1

РОЗРОБНИКИ ПРОГРАМИ:

доктор фізико-математичних наук, професор кафедри штучного інтелекту та програмного забезпечення **Куклін Володимир Михайлович;**

старший викладач кафедри штучного інтелекту та програмного забезпечення **Гущин Іван Валерійович;**

старший викладач кафедри штучного інтелекту та програмного забезпечення **Мішин Олександр Вікторович.**

Програму схвалено на засіданні кафедри штучного інтелекту та програмного забезпечення
Протокол від «29» серпня 2018 року № 1


Завідувач кафедри штучного інтелекту та програмного забезпечення


_____ (Куклін В.М.)

Програму погоджено методичною комісією факультету комп'ютерних наук

Протокол від «12» вересня 2018 року № 1

Голова методичної комісії факультету комп'ютерних наук


_____ (Васильєва Л.В.)



ВСТУП

Програма навчальної дисципліни «Розробка систем штучного інтелекту» складена відповідно до освітньо-професійної програми підготовки другого (магістерського) рівня вищої освіти за спеціальністю 122 «Комп'ютерні науки».

1. Опис навчальної дисципліни

1.1. Метою викладання навчальної дисципліни є формування у студентів теоретичних знань з предмету та практичних навичок у застосуванні методів машинного навчання у різноманітних класах задач та здобуття навичок вибору методів та інструментів, які найбільш придатні вирішувати поставлені задачі.

1.2. Основними завданнями вивчення дисципліни є:

- виявлення класів задач та підходів для їх вирішення
- ознайомлення з базовими програмними середовищами створення систем обробки та візуалізації даних..
- напрацювання студентами практичних навичок з роботи над задачами, які найбільш приближенні до реальних задач , пов'язаних з використанням технологій МН.
- ознайомлення з деякими головними бібліотеками та фрейморками для роботи над поставленою задачею.
- навчання продуктивної роботи у команді для виконання загальної задачі

1.3. Кількість кредитів – 10

1.4. Загальна кількість годин - 300

1.5. Характеристика навчальної дисципліни	
Нормативна / за вибором	
денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
1-й	1-й
Семестр	
1-й	1-й
Лекції	
32 год.	4 год.
Практичні, семінарські заняття	
16 год.	2 год.
Лабораторні заняття	
32 год.	2 год.
Самостійна робота	
180 год.	252 год.
Індивідуальні завдання	
40 год.	40 год.

1.6. Заплановані результати навчання

знати:

- теоретичні основи основних методів МН
- які саме методи МН можуть застосовуватися для окремого класу задач;
- декілька різних методів рішення окремих задач та їх відмінності.
- необхідність взагалі використовувати у окремій задачі той чи інший клас методів МН
- теоретичні основи видів навчання, вміти розпізнати вид навчання щодо окремої задачі;
- основні методи та інструментарій, що застосовується при розробці та реалізації систем МН

вміти:

- будувати комп'ютерні моделі МН на основі формулювання задач;
- обрати тип МН в залежності від класу поставленої задачі ;
- застосовувати методи оптимізації навчання, якщо це можливо та необхідно;
- аналізувати отримані результати та давати їх інтерпретацію.
- застосовувати отриманні теоретичні знання для обґрунтування статусу навчання власної моделі;
- самостійно розробляти архітектуру інтелектуальної системи, збирати та систематизувати дані та знання, необхідні для створення інтелектуальної системи;
- Самостійно або в команді реалізувати у програмному кодї протестований функціонал, який в підсумку оптимізує результат загальної моделі.

2. Тематичний план навчальної дисципліни

Розділ 1. Проектування агентних систем, які навчаються з підкріпленням

Тема 1. Знайомство з машинним навчанням

Що таке машинне навчання (МН). Місце МН у сучасному житті. Термінологія. Загальна структура типів задач, які можуть розв'язуватись за допомогою МН. Моделі навчання, такі як навчання з вчителем, без вчителя, змішанні. Місце навчання з підкріпленням у ієрархії моделей навчання. Приклади.

Тема 2. Вступ у навчання з підкріпленням.

Простір станів. Марковський процес прийняття рішень. Марківські стани. Вірогідності переходів між станами. Поняття агента та його валстивості. Миттева і відкладена нагороди . Простір дій агента. Поняття історій переміщень у просторі станів. Стан середовища та види середовища відносно агента.

Тема 3. Навчання з підкріпленням. Основні властивості середовищ та агентів

Повністю спостережувані середовища. Частково спостережувані середовища. Основні компоненти навчання з підкріпленням. Policy (лінія поведінки) агента. Функція вартості оцінки якості окремого стану. Модель прийняття рішень щодо окремого кроку. Види агентів. Прогнозування та контроль. Пошук оптимальної лінії поведінки агента.

Тема 4. Навчання з підкріпленням. Види навчання залежно від відомостей о середовищі та можливостей агента

Безмоделеве прогнозування та контроль. Апроксимація функції значень. SARSA. Q навчання. Deep Q навчання за допомогою нейронних мереж. Приклади

Тема 5. Огляд основних видів нейронних мереж (НМ).

НМ прямого поширення сигналу. НМ з наявністю зворотнього зв'язку. Глибокі НМ. Метода градієнтного спуску. Приклади використання окремих архітектур НМ.

Розділ 2. Лінійна регресійна модель

Тема 6. Регресійна модель. Основні поняття та визначення.

Поняття набору даних (data set), факторів та залежної змінної. Вивчення методу найменших квадратів та обчислення помилки моделі. Способи представлення набору даних в мові Python.

Тема 7. Порядок роботи с даними.

Знайомство з методами підготовки даних, доцільність їх використання. Аналіз факторів та аналіз їх впливу на залежну змінну. Основні прийоми роботи з набором даних в мові Python

Тема 8. Регресія однієї та декількох факторів (мультирегресія)

Розбір прикладів створення регресійної та мультирегресійної моделей. Основні регресії декількох факторів, поняття мультиколінеарності. Приклади створення різних регресійних моделей та візуалізації результатів в мові Python.

Розділ 3. Класифікаційні моделі.

Тема 9. Класифікація.

Визначення та терміни пов'язані з задачами класифікації. Логістична регресія.

Розбір типових класифікаційних задач

Тема 10. Алгоритми класифікації.

Знайомство з основними алгоритмами класифікації. Дерева рішень та метод найближчих сусідів.

Тема 11. Крос-валідація та вибір параметрів.

Визначення крос-валідації. Метод стримування. Кваліфікаційна перехресна перевірка. Leave-one-out перехресна перевірка. Зміщення та дисперсія.

Тема 12. Класифікаційна модель.

Оцінка якості класифікаційних моделей. ROC-криві, кількість помилково позитивних результатів, точність та чутливість.

Тема 13. Навчання без учителя.

Основні визначення. Кластеризація. Метод К-середніх.

3. Структура навчальної дисципліни

Назви розділів і тем	Кількість годин											
	денна форма						заочна форма					
	усього	у тому числі					усього	у тому числі				
		л	п	лаб	інд	с. р.		л	п	лаб	інд	с. р.
1	2	3	4	5	6	7	8	9	10	11	12	13
Розділ 1. Проектування агентних систем, які навчаються з підкріпленням.												
Тема 1. Знайомство с машинним навчанням.	18	2			1	15	19	1	1	1	1	15
Тема 2. Вступ у навчання з підкріпленням.	26	4	2		5	15	26	1			5	20
Тема 3. Навчання з підкріпленням. Основні властивості середовищ та агентів	34	4	2	8	5	15	32				5	27
Тема 4. Навчання з підкріпленням. Види навчання залежно від відомостей о середовищі та можливостей агента	32	4	2	8	3	15	30				3	27
Тема 5. Огляд основних видів нейронних мереж (НМ).	22	2	2		3	15	22				3	19
Разом за розділом 1	132	16	8	16	17	75	129	2	1	1	17	108
Розділ 2. Лінійна регресійна модель.												
Тема 6. Регресійна модель. Основні поняття та визначення.	20	2			3	15	20				3	17
Тема 7. Порядок роботи с даними.	22	2			5	15	22				5	17
Тема 8. Регресія однієї та декількох факторів(мультирегресія)	30	2	2	8	3	15	30				3	27
Разом за розділом 2	72	6	2	8	11	45	72				11	61
Розділ 3. Класифікаційні моделі.												
Тема 9. Класифікація	16	2	2		2	10	17	1	1	1	2	12
Тема 10. Алгоритми класифікації.	16	2	2		2	10	15	1			2	12
Тема 11. Крос-валідація та вибір параметрів.	16	2	2		2	10	16				2	14
Тема 12. Класифікаційна модель.	28	2		8	3	15	31				3	28
Тема 13. Навчання без учителя.	20	2			3	15	20				3	17
Разом за розділом 3	96	10	6	8	12	60	99	2	1	1	12	83
Усього годин	300	32	16	32	40	180	300	4	2	2	40	252

4. Теми практичних (лабораторних) занять

№ з/п	Назва теми	Кількість годин
Розділ 1. Проектування агентних систем, які навчаються з підкріпленням		
Практичні заняття		
1	«Ознайомлення з машинним навчанням (МН) ». Місто машинного навчання у сучасному житті. Приклади.	2
2	«Існуючі підходи до навчання МН моделей. Навчання з підкріпленням ». Приклади роботи подібних систем на Python.	2
3	«Навчання з підкріпленням». Приклади реалізації Марковських процесів прийняття рішень	2
4	«Навчання з підкріпленням». Приклади реалізації табличного навчання агентів.	2
Лабораторні заняття		
5	«Лабораторна робота №1». Використання табличного навчання агентів. <i>Table Q - learning</i>	8
6	«Лабораторна робота №2». Методи вирішення exploration-exploitation trade - off у задачах навчання з підкріпленням.	8
Розділ 2. Лінійна регресійна модель.		
Лабораторні заняття		
7	«Лабораторна робота №3». Створення регресійної моделі прогнозування затримки відправлення літаків.	6
Практичні заняття		
8	Робота з набором даних. Алгебраїчні дії з матрицями в Python.	
9	Доповнення неповних даних. Очищення даних.	1
10	Нормалізація та стандартизація. Категоризація, one hot encoding, label encoding, регуляризація.	1
11	Відображення результатів та збереження моделі.	
Розділ 3. Класифікаційні моделі.		
Практичні заняття		
12	Робота з текстовими даними у Python.	1
13	Векторизація набору даних. Розглядання різних видів векторизаторів.	1
14	Створення та навчання різних класифікаційних моделей.	2
15	Способи оцінки класифікаційної моделі	2
Лабораторні заняття		
16	«Лабораторна робота №4». Створення класифікаційної моделі для аналізу залежності розміру зарплати від текстового опис вакансії	10
	Усього	48

5. Завдання для самостійної роботи

№ з/п	Види, зміст самостійної роботи. Назва теми.		Кількість годин
Розділ 1. Проектування агентних систем, які навчаються з підкріпленням			
1	Опрацювання навчального матеріалу	Знайомство с машинним навчанням на прикладі розбору реалізації роботи з структурами даних у язику Python	15
2	Опрацювання навчального матеріалу	Реалізація зберігання стану агента у вигляді датафрейму, який має містити неповторювані елементи	15
3	Опрацювання навчального матеріалу	Реалізація “стратегії” агента, яка має становити з себе оптимізацію руху у просторі станів агента та вибір вірної дії на кожному кроці	15
4	Опрацювання навчального матеріалу	Огляд можливостей оптимізації алгоритму навчання за допомогою налаштування параметрів, таких як discount factor або learning rate	15
5	Опрацювання навчального матеріалу	Огляд прикладу роботи глибокої нейронної мережі для класифікації зображень з використанням бібліотек Theano або Tensorflow	15
Розділ 2. Лінійна регресійна модель			
6	Опрацювання навчального матеріалу	Завантаження набору даних з різних типів файлів та основні операції з наборами даних.	15
7	Опрацювання навчального матеріалу	Очищення даних. Знаходження кореляції у даних. Створення тренувальної та навчальної вибірок.	15
8	Опрацювання навчального матеріалу	Створення та ініціалізація регресійних моделей. Відображення та оцінка результатів роботи моделей. Вибір оптимальних параметрів.	15
Розділ 3. Класифікаційна модель			
9	Опрацювання навчального матеріалу	Вибір лише текстових ознак з набору даних та об’єднання їх у єдину ознаку.	10
10	Опрацювання навчального матеріалу	Векторизація датасету за допомогою різних векторизаторів та порівняння результатів.	10
11	Опрацювання навчального матеріалу	Створення різних класифікаційних моделей та візуалізація результатів.	10
12	Опрацювання навчального матеріалу	Вибір оптимальної класифікаційної моделі та її параметрів для підвищення точності.	15
13	Опрацювання навчального матеріалу	Аналіз результатів роботи моделі створенної під час роботи над лабораторною роботою №4.	15
		Виконання курсової роботи	40
	Усього		220

6. Індивідуальні завдання

Напрямки для виконання курсової роботи:

- Реалізація завдання класифікації методом logistic regression
 - Реалізація завдання класифікації методом SVM
 - Реалізація завдання класифікації методом decision tree
 - Реалізація завдання класифікації методом random forest
 - Реалізація завдання класифікації методом KNN
 - Реалізація завдання класифікації методом Naive Bayes
 - Реалізація завдання класифікації методом Neural Network
 - Використання бібліотеки scіru для вирішення завдання максимізації прибутку
 - Використання бібліотеки scіru для вирішення завдання мінімізації витрат
 - Огляд утиліти для розмітки зображень LabelImg
 - Огляд можливостей бібліотеки PyBrain
 - Огляд можливостей бібліотеки SimpleAI
 - Огляд можливостей бібліотеки Keras
 - Огляд можливостей бібліотеки Neurolab
 - Огляд можливостей бібліотеки Theano в роботі з глибокими нейронними мережами
 - Огляд можливостей бібліотеки Tensorflow в роботі з глибокими нейронними мережами
 - Огляд можливостей бібліотеки Brainforge
 - Sentiment analysis with NLTK and Naive Bayes
 - Огляд можливостей фреймворка глибинного навчання PyTorch
 - Реалізація завдання зменшення розмірності даних, використовуючи метод PCA бібліотеки sklearn
 - Рішення завдання "коктейльної вечірки" за допомогою методів бібліотеки sklearn (ICA)
 - Реалізація завдання кластеризації методом k-means бібліотеки sklearn
 - Використання Dropout підходу в задачах регуляризації
 - Огляд можливостей фреймворка Scab для створення рекомендаційних систем
 - Реалізація завдання регресії за допомогою ElasticNet
 - Реалізація CartPole завдання за допомогою нейронних мереж
 - Реалізація багат шарового перцептрона на основі функцій бібліотеки sklearn
 - Реалізація autoencoder
 - Огляд можливостей Open AI Universe
- 2 контрольні роботи

7. Методи контролю

Контроль засвоєння навчального матеріалу здійснюється шляхом:

- Проведення захисту курсових робіт
- Проведення іспиту з використанням інформаційних технологій, тобто на іспиті команди повинні реалізувати та захистити тестовий проект (у форматі “хакатону”),

захист якого повинен бути підкріплений звітом щодо внеску кожного учасника команди до результату.

- При виникненні спірної ситуації з оцінкою іспиту студентам будуть задані усні запитання з представленою нижче списку і у відповідності з отриманими відповідями буде виставлена оцінка на їх екзаменаційному звіті, де вони повинні будуть поставити свій підпис.

7.1 Додаткові усні питання до іспиту

1. Чим відрізняється регресія від класифікації
2. Які задачі машинного навчання ви знаєте
3. Що таке машинне навчання
4. Приклади роботи алгоритму навчання з вчителем
5. Приклади роботи алгоритму навчання без учителя
6. Приклади роботи алгоритму навчання з підкріпленням
7. Які існують методи оцінки результатів класифікаційної моделі
8. Які існують методи оцінки результатів регресійної моделі
9. Які існують етапи роботи з даними
10. Який зміст у негативній кореляції
11. Що таке мультиколінійність?
12. Які є підходи до вирішення проблеми мультиколінійності?
13. Що таке регуляризація?
14. Нормалізація та стандартизація. Що це таке та для чого використовується?
15. Який фізичний зміст у Q - таблиці
16. Що таке exploration-exploitation trade-off
17. Що таке Марківський стан
18. Вірогідності переходів між станами.
19. Поняття агента та його властивості.
20. Поняття нагороди. Миттєва і відкладена нагороди .
21. Простір дій агента.
22. Поняття історій переміщень у просторі станів.
23. Стан середовища та види середовища відносно агента.
24. Полісу (лінія поведінки) агента.
25. Функція вартості оцінки якості окремого стану.
26. Модель прийняття рішень щодо окремого кроку.
27. Види агентів
28. Прогнозування та контроль.

29. Пошук оптимальної лінії поведінки агента.

30. Що таке глибоке навчання за допомогою нейронних мереж

8. Схема нарахування балів

Розподіл балів для підсумкового семестрового контролю при проведенні семестрового екзамену.

Поточний контроль, самостійна робота, індивідуальні завдання															Екзамен		Сума		
Розділ 1					Розділ 2			Розділ 3					Контрольні роботи, передбачені навчальним		Індивідуальні завдання			Разом	
1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	КП	РР	60	40	100
10					5			5					5	5	20	10			

Шкала оцінювання

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка	
	для чотирирівневої шкали оцінювання	для дворівневої шкали оцінювання
90 – 100	відмінно	зараховано
70-89	добре	
50-69	задовільно	
1-49	незадовільно	не зараховано

Критерії оцінювання

Оцінка в балах	Оцінка за національною шкалою	Пояснення
90 – 100	Відмінно	Теоретичний зміст курсу освоєний цілком, необхідні практичні навички роботи з освоєним матеріалом сформовані, всі навчальні завдання, які передбачені програмою навчання виконані в повному обсязі, відмінна робота без помилок або з однією незначною помилкою.
70 – 89	Добре	Теоретичний зміст курсу освоєний цілком, практичні навички роботи з освоєним матеріалом в основному сформовані, всі навчальні завдання, які передбачені програмою навчання виконані, якість виконання жодного з них не оцінено мінімальним числом балів, деякі види завдань виконані з помилками, робота з декількома незначними помилками, або з однією – двома значними помилками.
50 – 69	Задовільно	Теоретичний зміст курсу освоєний не повністю, але прогалини не носять істотного характеру, необхідні практичні навички роботи з освоєним матеріалом в основному сформовані, більшість передбачених програмою навчання навчальних завдань виконано, деякі з виконаних завдань, містять помилки, робота з трьома значними помилками
1–49	Незадовільно	Теоретичний зміст курсу не освоєно, необхідні практичні навички роботи не сформовані, всі виконані навчальні

		завдання містять грубі помилки, додаткова самостійна робота над матеріалом курсу не приведе до значимого підвищення якості виконання навчальних завдань, робота, що потребує повної переробки
--	--	---

9. Рекомендована література

Основна література

1. Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction: [Електронний ресурс]. Copyright © 1993, 1994, Nikos Drakos, Computer Based Learning Unit, University of Leeds. URL: <https://pdfs.semanticscholar.org/aa32/c33e7c832e76040edc85e8922423b1a1db77.pdf>.
2. Sergey Levine , CS 294: Deep Reinforcement Learning, [Електронний ресурс]. URL: <http://rll.berkeley.edu/deeprlcourse/>
3. Denny Britz, Learning Reinforcement Learning (with Code, Exercises and Solutions), [Електронний ресурс]. 2016. URL: <http://www.wildml.com/2016/10/learning-reinforcement-learning/>
4. Abhijit Gosavi, A Tutorial for Reinforcement Learning, [Електронний ресурс]. URL: <https://web.mst.edu/~gosavia/tutorial.pdf>
5. Саймон Хайкин, «Нейронные сети полный курс», 2-е изд., испр.: пер. с англ. – М.:ООО «И.Д. Вильямс», 2006. 1104 с.
6. Осовский «Нейронные сети для обработки информации» Издательство: Финансы и статистика ISBN 5-279-02567-4, 83-7207-187-X; 2004 г.
7. Andrew Nj, Machine learning course notes, [Електронний ресурс]. URL: <http://cs229.stanford.edu/notes/>

10. Посилання на інформаційні ресурси в Інтернеті, відео-лекції, інше методичне забезпечення

1. Язык програмування з документацією, Python 3.x., [Електронний ресурс]. URL: www.python.org
2. Набір бібліотек для Python, Anaconda 3.x., [Електронний ресурс]. URL: www.continuum.io/ (includes Numpy, Matplotlib, sklearn, Ipython Notebook))
3. Набір інструментів Open AI Gym, [Електронний ресурс]. URL: <https://gym.openai.com>
4. Система командного спілкування, Slack, [Електронний ресурс]. URL: <https://slack.com/>
5. Репозиторії, Gitlab / Github / BitBucket, [Електронний ресурс]. URL: <https://gitlab.com/> , <https://github.com/> , <https://bitbucket.org/>
6. Приклади реалізації Reinforcement Learning алгоритмів, [Електронний ресурс]. URL: <https://github.com/rlcode/reinforcement-learning>
7. http://frnsys.com/ai_notes/artificial_intelligence/reinforcement_learning.html
8. Відео-лекції, Andrew Nj, Machine Learning, [Електронний ресурс]. URL: <https://www.youtube.com/playlist?list=PL0Smm0jPm9WcCsYvbhPCdizqNKps69W4Z>