

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ

Кафедра математичної інформатики



Кашпур О.Ф.

2019 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
МАШИННЕ НАВЧАННЯ
MACHINE LEARNING
для студентів / for students

галузь знань 12 – Інформаційні технології / Information Technologies

спеціальність 122 – Комп'ютерні науки / Computer Science

освітній рівень магістр / Master

освітня програма Штучний інтелект / Artificial Intelligence
вид дисципліни обов'язкова / mandatory

Форма навчання денна
Навчальний рік 2019/2020
Семестр 1
Кількість кредитів ECTS 5
Мова викладання, навчання та оцінювання англійська, українська /English, Ukrainian
Форма заключного контролю залік / credit

Викладачі: к.ф.-м.н., асистент Тарануха В.Ю. (лекції, практичні заняття)

Пролонговано: на 20__/20__ н.р. _____ (_____) «__» 20__ р.
(підпис, ПІБ, дата)

на 20__/20__ н.р. _____ (_____) «__» 20__ р.
(підпис, ПІБ, дата)

Розробник:

Терещенко Василь Миколайович, д. ф.-м. н., проф., завідувач кафедри математичної інформатики

Тарануха Володимир Юрійович, к.ф.-м.н., асистент кафедри математичної інформатики

ЗАТВЕРДЖЕНО

Завідувач кафедри математичної інформатики

В.М. Терещенко В.М.
(підпис) (прізвище та
ініціали)

Протокол № 10 від «23» травня 2019 р.

Схвалено науково-методичною комісією факультету комп'ютерних наук та кібернетики

Протокол від «30» серпня 2019 року № 1

Голова науково-методичної комісії Л.Л. (Омельчук Л.Л.)
(підпис) (прізвище та ініціали)

«30» серпня 2019 року

ВСТУП

1. Мета дисципліни – дати знання про теоретичні основи та практичного використання алгоритмів машинного навчання в тому числі з генетичних, еволюційних алгоритмів, інтелектуальних агентів та мультиагентних систем.

/

The purpose of the course - to provide knowledge about the theoretical foundations and practical use of machine learning algorithms, including genetic, evolutionary algorithms, intelligent agents and multi-agent systems.

2. Попередні вимоги до опанування або вибору навчальної дисципліни:

1. *Знати* деякі глави математичного аналізу, алгебри та теорії ймовірностей, дискретної математики, мати уявлення про методи розпізнавання. Знання технічної англійської мови на рівні B1.
2. *Вміти* використовувати сучасний математичний апарат, вміти будувати та використовувати нейронні мережі та прості методи машинного навчання, як то наївний Баєс, k-nn тощо.
3. *Володіти навичками* програмування мовою високого рівня, хоч би однією з: C++, C#, Java, Python.

/

Preliminary requirements to master or choosing of the course

1. *To know* certain chapters of mathematical analysis, linear algebra, probability theory and discrete mathematics, it is also desirable to have some understanding of the methods of pattern recognition. Level B1 technical English skills.
2. *To be able* to use a modern mathematical apparatus, be able to build and use neural networks and simple methods of machine learning, such as naive Bayes, k-nn, etc.
3. *To have* programming skills in high-level language, at least in one of: C++, C#, Java, Python.

3. Анотація навчальної дисципліни:

Навчальна дисципліна «Машинне навчання/Machine Learning» є складовою освітньо-наукової програми підготовки фахівців «Штучний інтелект» за другим (магістерським) рівнем вищої освіти *галузі знань* 12 «Інформаційні технології» *спеціальності* 122 «Комп'ютерні науки».

Дана дисципліна є обов'язковою навчальною дисципліною за програмою “Штучний інтелект”. Викладається у 1 семестрі 1 курсу магістратури в обсязі 5 кредитів ECTS.

У курсі передбачено 3 змістових частини, 2 контрольні роботи, 2 лабораторні роботи. Завершується дисципліна заліком в 1 семестрі 1 курсу магістратури.

/

Synopsis of the course:

The discipline "Machine Learning" is a component of the educational-scientific training program for the second (*master's*) level of higher education in the *field of knowledge* 12 "Information Technology" *specialty* 122 "Computer Science", educational-professional program "Artificial Intelligence".

This discipline is a mandatory discipline in the *program "Artificial Intelligence"*. It's taught in the 1st semester of the 1st year of master's degree in the amount of 5 ECTS credits. The course includes 3 modules, 2 module tests, 2 laboratory works. The discipline ends with a credit in the 1st semester of the 1st year of master's degree study.

4. Завдання (навчальні цілі):

Основними завданнями дисципліни «Машинне навчання» є набуття знань, умінь та навичок (компетентностей) на рівні новітніх досягнень в області машинного навчання відповідно до освітньої кваліфікації магістр комп'ютерних наук. Зокрема, розвивати:

- Здатність спілкуватися іноземною мовою (ЗК5).

- Здатність вирішувати складні задачі інтелектуальної обробки даних з використанням еволюційного моделювання, нейромережних технологій, застосування обчислювального інтелекту для розв'язання практичних задач в різних галузях професійної діяльності (СК8).
- Здатність розробляти та застосовувати індуктивні методи синтезу моделей, розпізнавання об'єктів на зображеннях, мультиагентні та нечіткі системи, нейромережі в процесі їх реалізації на сучасних високопродуктивних системах (СК9).

/

Learning objectives:

The main objectives of the discipline “Machine Learning” is acquisition of knowledge, skills and competencies at the level of the latest advances in computer science. In particular it aims to develop:

- Ability to communicate in a foreign language.
- Ability to solve complex problems of intellectual data processing using evolutionary modeling, neural network technologies, the use of computational intelligence to solve practical problems in various fields of professional activity.
- Ability to develop and apply inductive methods of model synthesis, object recognition in images, multiagent and fuzzy systems, neural networks in the process of their implementation on modern high-performance systems.

5. Результати навчання за дисципліною/ Results of learning:

Результат навчання (1. знати; 2. вміти; 3. комунікація; 4. автономність та відповідальність) Results of learning		Форми (та/або методи і технології) викладання і навчання/ Forms (and/or methods and technologies) of teaching	Методи оцінювання та пороговий критерій оцінювання (за необхідності)/ Methods of evaluation and evaluation threshold (if necessary)	Відсоток у підсумковій оцінці з дисципліни/ Percentage in final mark
Код	Результат навчання/ Learning results			
PH1.1	Знати основні поняття з машинного навчання, класи задач, межі придатності основних методів. / Know the basic concepts of machine learning, problem classes, the limits of suitability of basic methods.	Лекція, лабораторне заняття. / Lecture, practice	Активна робота на лекції, усні відповіді. Контрольна робота 1,2 (60% правильних відповідей) / <i>Active work on lectures, oral answers,</i> Test 1,2 (60% correct answers)	17%
PH 1.2	Знати генетичні та еволюційні алгоритми, баєсівські мережі, нейронні мережі, регресійний аналіз, мультиагентні системи, метод опорних векторів тощо. / Know genetic and evolutionary algorithms, Bayesian networks, neural networks, regression analysis, multiagent systems, support vector machines, etc.		18%	
PH 2.1	Вміти аналізувати практичні	Лабораторне	Лабораторна	10%

	задачі з метою визначення оптимального методу машинного навчання для їх розв'язання. / Be able to analyze practical problems in order to determine the optimal method of machine learning to solve them.	заняття, самостійна робота, / Practice, Individual work	робота 1,2, / Laboratory works 1,2	
PH 2.2	Вміти реалізувати відібрані шляхом аналізу задачі методи машинного навчання, вміти аналізувати причини отриманих результатів навчання та віднаходити способи покращення якості навчання. / To be able to implement the methods of machine learning selected by the analysis of the task, to be able to analyze the reasons for the obtained learning outcomes and to find ways to improve the quality of learning.			40%
PH 2.3	Вміти застосовувати програмні засоби розробки систем / Be able to use development tools			5%
PH 3.1	Обґрунтовувати власний погляд на задачу, спілкуватися з колегами з питань аналізу задач та проектування алгоритму / Be able to justify own view of the problem, communicate with colleagues in the design and development of programs, prepare written reports			5%
PH4.1	Відповідально ставитися до виконуваних робіт, нести відповідальність за їх якість / Responsibly treat the works performed, be responsible for their quality	Самостійна робота / Individual work	Виконання завдань, винесених на самостійну роботу / Accomplishment of tasks assigned to Individual work	5%

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання / Correspondence between learning results and program study results

Результати навчання дисципліни / Програмні результати навчання/	1.1	1.2	2.1	2.2	2.3	3.1	4.1
---	-----	-----	-----	-----	-----	-----	-----

Teaching results Program results of teaching							
<p>ПРН3. Опанувати нові інструменти роботи з даними, здійснюючи обробку веб-логів, текст-аналіз і машинне навчання, для прогнозування бізнес-процесів та ситуаційного управління, сентимент-аналізу відгуків, розробки рекомендаційних систем для сфери електронної комерції, медіа, соціальних мереж, банкінгу, реклами тощо.</p> <p>/</p> <p>Master new tools for working with data, web logs processing, text analysis and machine learning, for forecasting business processes and situational management, sentiment analysis of feedback, development of recommendation systems for e-commerce, media, social networks, banking, advertising etc.</p>	+		+	+	+		
<p>ПРН10. Використовувати інтелектуальні агенти, мультиагентні системи, машинне навчання та самонавчання, генетичні, кооперативні та розподілені еволюційні алгоритми для комп'ютерного розв'язання задач, що вимагають людського рівня мислення.</p> <p>/</p> <p>Use intelligent agents, multi-agent systems, machine learning and self-learning algorithms; genetic, cooperative and distributed evolutionary algorithms to solve computer problems that require a human level of thinking.</p>		+	+	+	+	+	+

7. Схема формування оцінки / Mark forming scheme.

7.1 Форми оцінювання студентів / Student evaluation forms:

- семестрове оцінювання/evaluation in semester:

Активна робота на лекції, усні відповіді/ Active work on lectures, oral answers: PH1.1, PH1.2 – 5 балів/3 бали;

2. Виконання завдань, винесених на самостійну роботу/ Tasks assigned to independent work: PH4.1 – 5 балів/3 бали

3. Контрольна робота 1/Test 1: PH 1.1 — 15 балів/9 балів.

4. Контрольна робота 2/Test 2: PH 1.1.,PH 1.2.— 15 балів/9 балів.

5. Лабораторна робота 1/Laboratory work 1 - Basic methods: PH2.1, PH 2.2, PH 2.3, PH3.1 – 25 балів/15 балів.

6. Лабораторна робота 2/Laboratory work 2 - Genetic Algorithms or Multiagent System: PH2.1, PH 2.2, PH 2.3,PH3.1 – 35 балів/21 бали.

- підсумкове оцінювання (у формі заліку)/final evaluation (credit):

- Залікові бали визначаються як сума оцінок/балів за всіма успішно оціненими результатами навчання, передбаченими даною програмою.

- Оцінки нижче від мінімального порогового рівня не додаються.

- Мінімальний пороговий рівень для сумарної оцінки за всіма компонентами становить 60% від максимально можливої кількості балів.

/

- Credit points are defined as the sum of grades / points for all successfully assessed learning outcomes provided by this program.

- Scores below the minimum threshold are not added.

- The minimum threshold for the total assessment of all components is 60% of the maximum possible number of points.

Умови лабораторних робіт/Laboratory works:

Лабораторна робота 1 Базові методи/Basic methods: Використовуючи відомий студенту метод машинного навчання спробувати побудувати модель процесу у певній предметній області: погоди, врожаїв, пробок на дорогах тощо./ Using the method of machine learning known to the student it's necessary to build a model of the process in a particular subject area: weather, harvests, traffic jams and more.

Лабораторна робота 2 Генетичний алгоритм або Мультиагентна система/Genetic Algorithms or Multiagent System: Для вибраної студентом предметної області реалізувати систему, що використовує генетичний алгоритм./ For the subject area chosen by the student it's necessary to implement a system that uses a genetic algorithm to solve a problem.

Запитання для контрольних робіт/Test questions

У першу контрольну роботу включається 1 теоретичне питання та 2 практичних завдання, за кожен пункт нараховується по 33,(3)% від кількості балів за контрольну./

First test includes 1 theoretical question and 2 practical tasks, each element contributes 33,(3)% of points.

Контрольна робота 1/Test 1:

1. Linear regression with one variable.
2. Cost function. Gradient descend.
3. Linear regression with multiple variables.
4. Logistic regression.
5. Regularization.
6. Evaluating a hypothesis model selection. Train/validation/test sets.
7. Bias–variance tradeoff.
8. Machine learning system design. Prioritizing and error analysis.

Контрольна робота 2/Test 2:

У другу контрольну включається 4 теоретичні питання. До переліку питань також включаються питання з першої контрольної роботи. За кожне питання нараховується по 25% від кількості балів за контрольну./

Second test includes 4 theoretical questions. They include also questions form first test. Each element contributes 25% of points.

9. K-means algorithm.
10. Neural networks. Multi-class classification.
11. Neural networks: learning. Cost function.
12. Backpropagation algorithm.
13. Support Vector Machines.

7.2 Організація оцінювання/Evaluation process:

Обов'язковим є виконання завдань, винесених на самостійну роботу, лабораторних робіт та контрольних робіт за графіком робочої програми./ It is mandatory to perform tasks assigned to independent work, laboratory work and tests according to the schedule of the work program.

Терміни проведення форм оцінювання/Deadlines:

1. Контрольна робота 1 (тест): до 5 тижня семестру./ Test1: up to the end of 5 weeks of the semester.

2. Контрольна робота 2 (тест): до 9 тижня семестру. / Test2: up to the end of 9 weeks of the semester.

3. Лабораторна робота 1 (проект): до 8 тижня семестру. / Laboratory work 1 (project): up to the end of 8 weeks of the semester.

4. Лабораторна робота 2 (проект): до 12 тижня семестру. / Laboratory work 2 (project): up to the end of 12 weeks of the semester.

Студент має право на однократне перескладання кожної контрольної роботи із можливістю отримання максимально 90% початково визначених за цю контрольну роботу балів. Термін перескладання визначається викладачем.

У випадку відсутності студента з поважних причин відпрацювання та перездачі контрольних робіт здійснюються у відповідності до „Положення про порядок оцінювання знань студентів при кредитно-модульній системі організації навчального процесу” від 1 жовтня 2010 року.

Студент має право здавати лабораторні роботи після закінчення визначеного для них терміну, але з втратою двох балів за кожен тиждень, який пройшов з моменту закінчення терміну її здачі.

/

The student can have second attempt to pass each test with (max. mark is 90% of initial course points for the test). The time of testing is set by lecturer.

Should the student be absent for valid reason re-testing and remedial training are performed according to “Regulations on evaluation of student’s knowledge under credit-module learning process” by 1 October 2010.

The student can submit lab works after deadline but with penalty to course earned points for each week after deadline.

7.3 Шкала відповідності оцінок / Mark correspondence scale

Зараховано / Passed	60-100
Не зараховано / Failed	0-59

8. Структура навчальної дисципліни. Тематичний план лекційних та лабораторних занять / The discipline structure. Lectures and Labs curriculum

№ п. п.	Назва лекції / Lecture name	Кількість годин / Hours		
		Лекції / Lectures	Лабораторні / Labs	Самост. робота / Individual work
Частина 1 Основи /Part 1 Basics				
1	<p>Тема 1. Вступ в машинне навчання. Лінійна регресія з однією змінною. Критерій якості. Градієнтний спуск. Елементи лінійної алгебри. Матриці та вектори. Операції.</p> <p><i>Самостійна робота:</i> побудувати лінійну регресію для заданих даних.</p> <p>Topic 1. Introduction. Linear regression with one variable. Cost function. Gradient descend. Linear algebra review. Matrices and vectors. Operations.</p> <p><i>Individual work:</i> construct a linear regression for the given data.</p>	2		8
2	<p>Тема 2. Лінійна регресія з кількома змінними. Градієнтний спуск для декількох змінних. Масштабування ознак. Швидкість навчання. Ознаки та поліноміальна регресія.</p> <p><i>Самостійна робота:</i> побудувати поліноміальну регресію для заданих даних.</p> <p>Topic 2. Linear regression with multiple variables. Gradient descent for multiple variables. Feature scaling. Learning rate. Features and polynomial regression.</p> <p><i>Individual work:</i> construct a polynomial regression for the given data.</p>	2	2	8
3	<p>Тема 3. Логістична регресія. Класифікація. Представлення гіпотези. Границя рішення. Регуляризація. Перенавчання. Регуляризована лінійна та логістична регресія.</p> <p><i>Самостійна робота:</i> побудувати логістичну регресію для заданих даних.</p> <p>Topic 3. Logistic regression. Classification. Hypothesis representation. Decision boundary. Regularization. Overfitting. Regularized linear and logistic regression.</p> <p><i>Individual work:</i> construct a logistic regression for the given data.</p>	2		8
4	<p>Тема 4. Застосування машинного навчання. Оцінка вибору моделей гіпотез та навчальна / валідаційна / тестова вибірка. Компроміс зсуву та дисперсії. Дизайн системи машинного навчання. Визначення пріоритетів роботи. Аналіз помилок.</p> <p><i>Самостійна робота:</i> спроектувати систему машинного навчання для предметної області за</p>	1	2	8

	<p>вибором студента.</p> <p>Topic 4. Applying machine learning. Evaluating a hypothesis model selection and train/ validation/ test sets. Bias–variance tradeoff. Machine learning system design. Prioritizing what to work on. Error analysis.</p> <p><i>Individual work:</i> design a machine learning system for the subject field of the student's choice.</p>			
	Контрольна робота № 1/ Test 1.	1		
Частина 2 Нейромережі та інше / Part2 Neural Networks and other methods				
5	<p>Тема 5. Навчання без учителя. Алгоритм к-середніх. Алгоритм с-середніх. Випадкова ініціалізація. Вибір числа кластерів. Зменшення розміру задачі: стиснення даних та візуалізація.</p> <p><i>Самостійна робота:</i> для заданих даних реалізувати к-середніх, візуалізувати результати.</p> <p>Topic 5. Unsupervised learning. K-means algorithm. C-means algorithm. Random initialization. Choosing the number of clusters. Dimensionality reduction: data compression and visualization.</p> <p><i>Individual work:</i> for the given data one needs to execute k-means and to visualize results.</p>	2		8
6	<p>Тема 6. Метод опорних векторів. Ядра.</p> <p><i>Самостійна робота:</i> для заданих даних виконати метод опорних векторів.</p> <p>Topic 6. Support Vector Machines. Kernels.</p> <p><i>Individual work:</i> for the given data one needs to execute SVM.</p>	2	2	8
7	<p>Тема 7. Нейронні мережі: навчання. Критерій якості. Алгоритм зворотного розповсюдження помилки.</p> <p><i>Самостійна робота:</i> для заданих даних навчити перцептрон.</p> <p>Topic 7. Neural networks: learning. Cost function Backpropagation algorithm.</p> <p><i>Individual work:</i> to train a perceptron for the given data.</p>	2		9
8	<p>Тема 7 (продовження). Нейронні мережі. Класифікація на багато класів.</p> <p><i>Самостійна робота:</i> для заданих даних навчити згорткову нейромережу.</p> <p>Topic 7 (cont.) Neural networks. Multi-class classification.</p> <p><i>Individual work:</i> for the given data to train a convolutional neural network.</p>	1	2	9
	Контрольна робота № 2/ Test 2.	1		
Частина 3 Великі дані / Part 3 Big Data				
9	<p>Тема 8. Генетичні алгоритми. Еволюційні алгоритми. Боротьба з застряганням у локальному оптимумі.</p> <p><i>Самостійна робота:</i> для вибраної студентом предметної області розв'язати задачу методом</p>	2		9

	генетичних алгоритмів. Topic 8. Genetic algorithms. Evolutionary algorithms. Fighting local optimum pit. <i>Individual work:</i> for the subject field chosen by the student to solve the problem by genetic algorithms.			
10	Тема 9. Мультиагентне навчання. <i>Самостійна робота:</i> у предметній області вибраній для генетичних алгоритмів описати процедуру, переваги та недоліки мультиагентного навчання. Topic 9. Multiagent learning. <i>Individual work:</i> describe the procedure, advantages and disadvantages of multi-agent training in the subject area selected for genetic algorithms.	2	2	9
11	Тема 10. Стохастичний градієнтний спуск. Пакетний градієнтний спуск з малим охоптом. Сходження стохастичного градієнтного спуску <i>Самостійна робота:</i> для заданих даних перевірити ефективність стохастичного градієнтного спуску при різних значеннях параметрів алгоритму. Topic 10. Stochastic gradient descent. Mini-batch gradient descent. Stochastic gradient descent convergence. <i>Individual work:</i> for the given data to check efficiency of stochastic gradient descent at various values of parameters of algorithm	2		9
12	Тема 11. Виявлення аномалій. Гауссовий розподіл і алгоритм для нього. Багатовимірний гауссовий розподіл. Рекомендаційні системи. Рекомендації на основі вмісту. Колаборативна фільтрація. <i>Самостійна робота:</i> виявити аномалії у заданому наборі даних. Topic 11. Anomaly Detection. Gaussian Distribution and algorithm. Multivariate Gaussian Distribution. Recommender Systems. Content Based Recommendations. Collaborative Filtering. <i>Individual work:</i> detect anomalies in a given data set.	2	2	9
13	Тема 12. Машинне навчання у великих задачах. map-reduce та паралелізм даних. Навчання з великими наборами даних. <i>Самостійна робота:</i> для даних з лекції 11 реалізувати map-reduce. Topic 12. Large scale machine learning. Map-reduce and data parallelism. Learning with large datasets. <i>Individual work:</i> for the data from lecture 11 to implement map-reduce algorithm.	2		8
	ВСЬОГО/ TOTAL	26	12	110

Загальний обсяг 150 год., в тому числі/ **Total duration** 150 hours, namely:
Лекції/ Lectures – 26 год./h.

Лабораторні роботи / Laboratory works – 12 год./h.

Консультації/ Consultations - 2 год./h.

Самостійна робота/ Individual work - 110 год. /h.

9. Рекомендовані джерела/Literature:

Основні / Main:

1. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989.
2. Флах П. Машинное обучение. — М.: ДМК Пресс, 2015. — 400 с
3. Уоссермен, Ф. Нейрокомпьютерная техника: Теория и практика — М.: Мир, 1992. — 240 с
4. Semi-Supervised Learning Edited by Olivier Chapelle Bernhard Schölkopf and Alexander Zien MIT 2006, 528 pp.
5. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Множественная регрессия — 3-е изд. — М.: Диалектика, 2007. — С. 912.
6. Martin T Hagan Neural Network Design (2nd Edition) / Martin T Hagan, Howard B Demuth, Mark H Beale, Orlando De Jesús // Campus Pub. Service, University of Colorado Bookstore, 2002, 736 pp.
7. Stone, James V.. Independent component analysis: a tutorial introduction. Cambridge, Massachusetts: MIT Press. 2004 200 pp.
8. Емельянов В. В., Курейчик В. В., Курейчик В. М. Теория и практика эволюционного моделирования. — М: Физматлит, 2003. — 432 с.

В тому числі інтернет-ресурси / Internet resources:

- Suvrit Sra 2005 Generalized Nonnegative Matrix Approximations with Bregman Divergences Advances in Neural Information Processing Systems 18 (NIPS 2005)
<http://papers.nips.cc/book/advances-in-neural-information-processing-systems-18-2005>

Додаткові / Additional:

1. Шлезингер М., Главач В. Десять лекций по статистическому и структурному распознаванию. — Киев: Наукова думка, 2004.
2. Finn V. Jensen Thomas D. Nielsen Bayesian Networks and Decision Graphs Springer, New York, NY 2001, 447 pp.

В тому числі інтернет-ресурси / Internet resources:

3. Caruana, R., 1998. Multitask learning. In Learning to learn (pp. 95-133). Springer, Boston, MA.
<http://www.cs.cornell.edu/~caruana/mlj97.pdf>