

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ТАРАСА ШЕВЧЕНКА

МЕХАНІКО-МАТЕМАТИЧНИЙ ФАКУЛЬТЕТ

Кафедра теорії ймовірностей, статистики та актуарної математики



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Алгоритми машинного навчання  
для студентів

галузь знань	11 «Математика та статистика»
спеціальність	112 «Статистика»
освітній рівень	перший (бакалавр)
освітня програма	«Статистика»
вид дисципліни	вибіркова

Форма навчання	денна
Навчальний рік	2020/2021
Семестр	6
Кількість кредитів ECTS	4
Мова викладання, навчання та оцінювання	українська
Форма заключного контролю	іспит

Викладачі: доцент Голомозий В. В., канд.ф.-м.н., доцент кафедри теорії ймовірностей та актуарної математики

Пролонговано: на 20 /20 н.р. ( ) « » 20 р.  
на 20 /20 н.р. ( ) « » 20 р.

КИЇВ – 2020

Розробник: Голомозий В.В., к.ф.м.н., доцент кафедри теорії ймовірностей, статистики та актуарної математики.

ЗАТВЕДЖЕНО  
Зав. кафедри  
кафедри теорії ймовірностей,  
статистики та актуарної математики  
Мішура Ю.С.

Протокол № 1 від 28.08.2020 р.

Схвалено науково-методичною комісією механіко-математичного факультету

Протокол від "31" 08 2020 року № 1

Голова науково-методичної комісії Олійник А.С. професор, д.ф.-м.н. Олійник А.С.  
(підпис)

**1. Мета дисципліни** – оволодіння сучасними методами, теоретичними положеннями та основними застосуваннями алгоритмів машинного навчання, зокрема алгоритмів регресії та класифікації.

**2. Попередні вимоги до опанування або вибору навчальної дисципліни:**

1. *Знати:* основи математичного аналізу, алгебри, аналітичної геометрії та теорії ймовірностей.
2. *Вміти:* знаходити екстремуми гладких функцій, обчислювати математичні сподівання.
3. *Володіти елементарними навичками:* робота з даними, графічний аналіз даних.

**3. Анотація навчальної дисципліни:**

Навчальна дисципліна «Алгоритми машинного навчання» є складовою освітньої програми підготовки фахівців за освітнім рівнем «бакалавр» галузі знань 11 Математика та статистика зі спеціальності 112 Статистика освітньої програми «Статистика».

Дана дисципліна є вибірковою. Дисципліна «Алгоритми машинного навчання» вивчає методи класифікації та регресійного аналізу. Зокрема, обчислення оцінки найменших квадратів, оцінки гребеневої регресії, використання методу k-найближчих сусідів, логістичної регресії, опорної машини векторів, алгоритмів заснованих на деревах, алгоритмів навчання без учителя та щільних нейронних мереж.

Викладається у 6 семестрі в обсязі 120 год. (*4 кредити ECTS<sup>1</sup>*) зокрема: *лекції – всього 28 год, лабораторії 28 год., самостійна робота – 64 год.* У курсі передбачено 2 змістових модулів, колоквиум та 4 лабораторних роботи. Завершується дисципліна **іспитом у шостому семестрі.**

**4. Завдання (навчальні цілі):** формування здатності розв'язувати складні спеціалізовані математичні та статистичні задачі, що характеризується комплексністю і невизначеністю умов і передбачає застосування теоретико-ймовірнісних і статистичних методів; набуття знань, умінь та навичок (компетентностей) на рівні новітніх досягнень у математиці та статистиці, відповідно до освітнього рівня «Бакалавр». Зокрема, професійне оволодіння компетентностями:

- 1) Здатність до абстрактного мислення, аналізу та синтезу (ЗК-3).
- 2) Здатність застосовувати знання у практичних ситуаціях (ЗК-4).
- 3) Знання й розуміння предметної області та професійної діяльності (ЗК-5).
- 4) Здатність спілкуватися українською мовою як усно, так і письмово (ЗК-6).
- 5) Здатність вчитися і оволодівати сучасними знаннями (ЗК-9).
- 6) Здатність до пошуку, обробки та аналізу інформації з різних джерел (ЗК-10).
- 7) Здатність приймати обґрунтовані рішення (ЗК-11).
- 8) Здатність працювати автономно (ЗК-14).
- 9) Визначеність і наполегливість щодо поставлених завдань і взятих обов'язків. (ЗК-15).
- 10) Здатність оцінювати та забезпечувати якість виконуваних робіт (ЗК-16).
- 11) Здатність діяти на основі етичних міркувань (мотивів). (ЗК-17).
- 12) Здатність здійснювати логічні математичні міркування із чітким зазначенням припущень та висновків (СК-3).

<sup>1</sup> кредитів ECTS – кредит кратний 30 годинам.

- 13) Здатність до математичного формулювання задач та вибору методів їх розв'язання (СК-4).
- 14) Здатність до кількісно-статистичного мислення (СК-5).
- 15) Здатність до ймовірнісного мислення, що передбачає сприйняття стохастичної природи явищ (СК-6).
- 16) Здатність робити якісні висновки з кількісних даних (СК-7).
- 17) Здатність проводити дослідження ймовірнісно-статистичних моделей та інтерпретувати одержані результати (СК-10).
- 18) Здатність подавати статистичні процедури та результати їхнього застосування у формі, придатній для цільової аудиторії, до якої звертаються, як усно, так і письмово (СК-13).
- 19) Здатність до аналізу основ і властивостей статистичних алгоритмів та розуміння переваг тих чи інших підходів, у тому числі до оцінки їх обґрунтованості й ефективності (СК-14).

## 5. Результати навчання за дисципліною:

Результат навчання (РН) (1. знати; 2. вміти; 3. комунікація; 4. автономність та відповідальність)		Форми (та/або методи і технології) викладання і навчання	Методи оцінювання та пороговий критерій оцінювання (за необхідності)	Відсоток у підсумковій оцінці з дисципліни
Код	Результат навчання (Формуються розробником)			
РН 1.1	Знати формулу для оцінки найменших квадратів та оцінки гребеневої регресії.	<i>Лекція, лаборатор на робота</i>	<i>Іспит, активна робота на лекції, усні відповіді, колоквиум</i>	5%
РН 1.2	Знати принципи роботи логістичної регресії та методу k-найближчих сусідів.			5%
РН 1.3	Знати як будується класифікатор для опорної машини векторів, та в чому полягає ядерне перетворення.			10%
РН 1.4	Знати основні методи побудови алгоритмів заснованих на деревах.			10%
РН 1.5	Знати як влаштована, та які задачі можна розв'язувати за допомогою щільної нейронної мережі.			10%
РН 2.1	Вміти обчислювати оцінку найменших квадратів та гребеневої регресії	<i>Лаборатор на робота, Лаборатор на робота</i>	<i>Контрольна робота 1 (60% правильних відповідей), виконання лабораторних робіт, виконання завдань, винесених на самостійну роботу</i>	15%
РН 2.2	Вміти використовувати основні методи класифікації, такі як: логістична регресія, опорна машина векторів, класифікаційні дерева та методи анасмбування засновані на деревах. Вміти використовувати алгоритми кластеризації.			15%
РН 2.3	Вміти обчислювати коефіцієнти щільної нейронної мережі та використовувати щільні нейронні мережі у задачах регресії та класифікації.			25%
РН 3.1	Здатність грамотно будувати комунікацію,	<i>Активна</i>	<i>Виконання</i>	2.5%

	виходячи з мети і ситуації спілкування	<i>робота на лекції, Лаборатор на робота</i>	<i>лабораторних робіт, усні відповіді</i>	
PH 3.2	Вироблення навиків командної роботи	<i>Лаборатор на робота</i>	<i>Виконання лабораторних робіт, усні відповіді</i>	2.5%

## 6. Співвідношення результатів навчання дисципліни із програмними результатами навчання

Результати навчання дисципліни	PH 1.1	PH 1.2	PH 1.3	PH 1.4	PH 1.5	PH 2.1	PH 2.2	PH 2.3	PH 3.1	PH 3.2
<b>Програмні результати навчання</b>										
<b>PH-1</b> - Здійснювати професійну письмову й усну комунікацію українською мовою та, принаймні, однією з іноземних мов	+	+	+	+	+	+	+	+	+	+
<b>PH-6</b> - Володіти знаннями та вміннями з імовірнісних і статистичних розділів математики: побудова ймовірнісних просторів, обчислення ймовірностей подій та характеристик випадкових величин і векторів, граничні теореми, характеристики випадкових процесів, оцінювання характеристик сукупностей на основі спостережень, формулювання та перевірка статистичних гіпотез	+	+	+	+	+	+	+	+	+	+
<b>PH-8</b> - Вміти працювати з різними типами збіжності випадкових величин та розподілів, користуватися граничними законами теорії ймовірностей	+	+	+	+	+	+	+	+	+	+
<b>PH-9</b> - Вміти визначати числові та якісні характеристики випадкових подій, величин, елементів, процесів	+	+	+	+	+	+	+	+	+	+
<b>PH-10</b> - Вміти здійснювати статистичне точкове, інтервальне оцінювання параметрів розподілів випадкових величин і процесів, непараметричне оцінювання, тестувати статистичні гіпотези	+	+	+	+	+	+	+	+	+	+
<b>PH-11</b> - Вміти аналізувати та прогнозувати лінійні статистичні моделі та моделі регресії, оцінювати їхні параметри	+	+					+	+		
<b>PH-12</b> - Вміти збирати та обробляти дані, застосовувати статистичні процедури для аналізу даних за допомогою обчислювальної техніки та програмних засобів	+	+	+	+	+	+	+	+	+	+
<b>PH-13</b> - Вміти моделювати реалізації випадкових величин і процесів та	+	+	+	+	+	+	+	+	+	+

використовувати результати моделювання для верифікації й аналізування ефективності статистичних процедур										
<b>PH-15</b> - Володіти математичними та статистичними методами аналізу, прогнозування та оцінки параметрів математичних моделей, статистичними методами інтерпретації та обробки числових даних	+	+	+	+	+	+	+	+	+	+
<b>PH-16</b> - Вміти використовувати в практичній діяльності спеціалізоване статистичне програмне забезпечення	+	+	+	+	+	+	+			
<b>PH-17</b> - Знати методи моделювання природничих та/або соціальних процесів	+	+	+	+	+	+	+	+	+	+

## 7. Схема формування оцінки.

### 7.1. Форми оцінювання студентів:

#### - оцінювання впродовж навчального періоду:

1. Активна робота на лекції, усні відповіді: PH1.1,PH1.2,PH1.3, PH1.4, PH1.5, PH2.1, PH2.2, PH2.3, PH3.1, PH3.2 – 5 балів/3 бали;
2. Виконання завдань, винесених на самостійну роботу: PH2.1, PH2.2 – 5 балів/3 бали;
3. Колоквіум: PH1.1, PH1.2, PH1.3, PH1.4 – 20 балів/12 балів
4. Виконання лабораторних робіт: PH2.1, PH2.2, PH2.3, PH3.1, PH3.2 – 30 балів/17 балів;  
Разом має бути 60/35

#### - підсумкове оцінювання: іспит.

- максимальна кількість балів, які можуть бути отримані: 40 балів;
- результати навчання, які будуть оцінюватись: PH1.1, PH1.2, PH1.3, PH1.4, PH1.5, PH2.1, PH2.2, PH2.3
- форма проведення і види завдань: письмова робота.

### 7.2. Організація оцінювання:

Критично-розрахунковий мінімум балів за навчання впродовж семестру становить **20** балів, рекомендований мінімум, розрахований з урахуванням специфіки дисципліни становить **35** балів. Студенти, які протягом семестру набрали сумарно меншу кількість балів ніж рекомендований мінімум **35** балів для підвищення балів отримують можливість написати додаткову контрольну роботу та доскласти домашні завдання. Мінімальна кількість балів, які додаються до семестрових – 24 бали, тобто, якщо оцінка студента на іспиті є нижчою від мінімального порогового рівня (24 бали), то бали за іспит не додаються до семестрової оцінки (вважаються рівними нулю), а підсумкова оцінка із дисципліни є незадовільною;

У випадку відсутності студента з поважних причин відпрацювання та перездачі форм контролю здійснюються у відповідності до „Положення про організацію освітнього процесу в Київському національному університеті імені Тараса Шевченка” (2018), <http://www.univ.kiev.ua/pdfs/official/Organization-of-the-educational-process.pdf>.

Форма іспиту в кожному семестрі – письмово-усна. Іспит складається з тесту та екзаменаційного білету. Тест складається з 20 запитань. Кожне запитання оцінюється в 1 бал. Екзаменаційний білет

іспиту складається із 2 завдань, перше є теоретичними, друге – задача. Задача оцінюється в 10 балів, теоретичне питання в 5 балів. Додатково від 0 до 5 балів студент отримує за усне опитування. Всього за іспит можна отримати від 0 до 40 балів.

**Терміни проведення форм оцінювання:**

1. Колоквіум: на 10-му тижні навчального періоду.

**7.3. Шкала відповідності оцінок**

<b>Відмінно / Excellent</b>	90-100
<b>Добре / Good</b>	75-89
<b>Задовільно / Satisfactory</b>	60-74
<b>Незадовільно / Fail</b>	0-59

**8. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ. ТЕМАТИЧНИЙ ПЛАН ЛЕКЦІЙ І ПРАКТИЧНИХ ЗАНЯТЬ**

теми	Назва теми	Кількість годин				
		Лекції	Лабораторні роботи	Самост. робота	Контр. модульна робота	Інші форми контролю
<b>Змістовий модуль 1 „Базові алгоритми”</b>						
1	Регресійний аналіз, гребенева регресія.	4	2	10		
2	Метод k-наближчих сусідів та логістична регресія.	4	4	10		
3	Опорна машина векторів.	4	8	10		
<b>Змістовий модуль 2 „Ієрархічні алгоритми”</b>						
1	Дерева прийняття рішень та регресійні дерева. Алгоритми ансамблювання.	4	4	10		
2	Алгоритми навчання без учителя.	6	6	10		2
3	Щільні нейронні мережі.	8	6	10		
Всього годин за III семестр		28	28	64		2

**Загальний обсяг 120 годин, у тому числі:  
лекції – 28 годин,  
практичні заняття – 28 годин,  
самостійна робота – 64 годин.**

## 9. Рекомендовані джерела

### Основні:

1. J. Faraway, Practical Regression and Anova using R. – <https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>, 2002
2. T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning. – Springer Series in Statistics, 2008.- 745 с.
3. G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R. Springer Texts in Statistics, 434 p.
4. C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006, 749p.

### Додаткові:

1. K. Murphy, Machine Learning: A Probabilistic Perspective – The MIT Press, 2012, 1098p.
2. P. Norvig, S. Russel, Artificial Intelligence: A Modern Approach – Prentice Hall, 2020, 1136p.
3. Z. Zhou, Ensemble Methods. Foundations and Algorithms. – Chapman & Hall/CRC, 2012, 234p.