

Міністерство освіти і науки України
Харківський національний університет імені В.Н. Каразіна
Кафедра прикладної математики

“ЗАТВЕРДЖУЮ”

В.о. декана факультету
математики і інформатики _____

Світлана МЕНЯЙЛОВА _____



серпень 2025 р.

РОБОЧА ПРОГРАМА НАЧАЛЬНОЇ ДИСЦИПЛІНИ

Алгоритми аналізу даних та штучні нейронні мережі

рівень вищої освіти _____ магістр _____

галузь знань _____ F Інформаційні технології _____

спеціальність _____ F1 Прикладна математика _____

освітня програма _____ Прикладна математика _____

спеціалізація _____

вид дисципліни _____ обов'язкова _____

факультет _____ математики і інформатики _____

2025 / 2026 навчальний рік

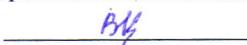
Програму рекомендовано до затвердження вченою радою факультету математики і інформатики
 “26” серпня 2025 року, протокол № 10

РОЗРОБНИКИ ПРОГРАМИ:

Степанова Катерина Вадимівна, кандидат фізико-математичних наук, доцент, доцент кафедри прикладної математики

Карєва Валерія Віталіївна, доктор філософії з Прикладної математики, доцент кафедри прикладної математики

Програму схвалено на засіданні кафедри Прикладної математики
 Протокол від “26” серпня 2025 року № 13

Завідувач кафедри Прикладної математики
 Валерій КОРОБОВ

Програму погоджено з гарантом освітньо-професійної Прикладна математика

Гарант освітньо-професійної програми Прикладна математика
 Олексій ШВЕНЬ

Програму погоджено з гарантом освітньо-наукової Прикладна математика

Гарант освітньо-наукової програми Прикладна математика
 Світлана ІГНАТОВИЧ

Програму погоджено науково-методичною комісією факультету математики і інформатики Харківського національного університету імені В.Н. Каразіна

Протокол від “26” серпня 2025 року, протокол № 1

Голова науково-методичної комісії факультету математики і інформатики
 Євген МЕНЯЙЛОВ

ВСТУП

Програма навчальної дисципліни “**Алгоритми аналізу даних та штучні нейронні мережі**” складена відповідно до освітньо-професійної і освітньо-наукової програм підготовки магістра спеціальності F1 Прикладна математика.

1. Опис навчальної дисципліни

1.1. Метою викладання навчальної дисципліни є надання майбутнім спеціалістам теоретичних знань та практичних навичок з основ розробки і аналізу алгоритмів для розв’язання різних прикладних задач.

1.2. Основні завдання вивчення дисципліни:

- ознайомити студентів із основними поняттями науки даних, штучних нейронних мереж;

- ознайомлення з основними методами аналізу алгоритмів;

- ознайомлення з деякими найбільш важливими алгоритмами;

- створення ефективних алгоритмів при розв’язанні прикладних задач.

1.3. Кількість кредитів - 6

1.4. Загальна кількість годин - 180

1.5. Характеристика навчальної дисципліни	
Обов’язкова	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
1-й	
Семестр	
1-й	
Лекції	
32 год.	
Практичні, семінарські заняття	
32 год.	
Лабораторні заняття	
год.	
Самостійна робота	
116 год.	
у тому числі індивідуальні завдання	
10 год.	

* у разі формування малочисельних груп обсяг аудиторного навчального навантаження, відведеного на вивчення навчальної дисципліни, зменшується відповідно до Положення про планування й звітування науково-педагогічних працівників Харківського національного університету імені В.Н. Каразіна.

1.6. Перелік компетентностей, що формує дана дисципліна

1.6.1. Формування наступних інтегральної та загальних компетентностей:

ІК01. Здатність розв'язувати складні задачі і проблеми у професійної діяльності з прикладної математики та/або у процесі навчання, що передбачає проведення досліджень та/або здійснення інновацій та характеризується невизначеністю умов і вимог.

ЗК02. Здатність застосовувати знання у практичних ситуаціях.

ЗК03. Знання та розуміння предметної області та розуміння професійної діяльності.

ЗК05. Здатність вчитися і оволодівати сучасними знаннями.

1.6.2. Формування наступних фахових компетентностей:

ФК02. Уявлення про прикладні задачі, які можуть бути досліджені за допомогою сучасних математичних методів, здатність до розуміння методів побудови і якісного і кількісного аналізу математичних моделей природних, техногенних, економічних та соціальних об'єктів та процесів.

ФК03. Здатність скористатися існуючими програмними засобами для проведення обчислень, пошуку інформації, оформлення результатів роботи тощо.

ФК04. Здатність аналізувати, вдосконалювати і створювати нові математичні моделі систем і процесів, аналізувати межі застосовності моделей.

ФК05. Здатність висувати нові гіпотези щодо поведінки математичної моделі, перевіряти їх справедливість у ході аналітичного дослідження моделі або чисельного експерименту, систематизувати отримані результати, застосовувати математичний апарат для доведення або спростування висунутих гіпотез, досліджувати межі застосування отриманих результатів.

ФК06. Здатність удосконалити існуючі і розробити нові алгоритми для дослідження та розв'язання різноманітних задач, що виникають при математичному моделюванні та аналізі даних.

ФК13. Здатність розуміти і застосовувати методи сучасного аналізу даних до розв'язання різноманітних прикладних задач.

1.7. Перелік результатів навчання, що формує дана дисципліна

Згідно з вимогами освітньо-професійної програми студенти повинні досягти таких програмних результатів навчання:

РН02. Уміти формалізувати прикладні задачі, формулювати їх математичну постановку, обирати раціональний метод їх розв'язання, оцінювати адекватність отриманих результатів, аналізувати їх взаємозв'язок з раніше відомими результатами і їх практичну значимість.

РН04. Володіти методами розробки, якісного та кількісного аналізу математичних моделей об'єктів та процесів, дослідження і використання цих моделей.

РН05. Володіти сучасними методами аналізу даних, вміти застосовувати їх для розв'язання прикладних задач.

РН06. Вміти застосовувати сучасні технології програмування та розроблення програмного забезпечення для розв'язання прикладних задач, у тому числі для проведення чисельних експериментів.

РН12. Вміти здійснювати науково-технічний пошук у сучасних джерелах інформації, уникаючи при цьому академічної недоброчесності, аналізувати і співвідносити результати з

різних джерел, орієнтуватися у новітніх наукових напрямках і їх застосуваннях, оформити результати досліджень у вигляді завершеної наукової роботи.

1.8. Пререквізити: вказати перелік дисциплін, що передують вивченню даної дисципліни
Курси з математичного аналізу, лінійної алгебри, теорії ймовірностей і програмування на попередньому рівні освіти

2. Тематичний план навчальної дисципліни

Тема 1. Вступ до алгоритмів науки даних. Логістична та гребнева регресія.

Основні поняття науки даних. Схема базових алгоритмів для науки даних та сфера їх застосування. Регресія. Гребнева регресія. Алгоритм логістичної регресії. Передбачення достовірностей подій.

Тема 2. Лінійна регресія.

Прогнозування довільних значень за допомогою лінійної регресії. Оцінка якості моделі за допомогою R2 Score.

Тема 3. Мультилінійна регресія.

Розподіл даних щодо вибірки на тренувальну множину та тестову множину. Означення крос-валідації. Визначення доцільних факторів для мультилінійної регресії та застосування цього алгоритму для прикладних задач.

Тема 4. Класифікація.

Класифікація даних за допомогою алгоритму дерева рішень. Візуалізація побудови рішень на базі впливових факторів за допомогою graphviz. Обґрунтування отриманих результатів за допомогою confusion matrix.

Тема 5. Кластеризація.

Застосування алгоритму KMeans, що кластеризує дані. Відокремлення зразків в n групах рівної дисперсії. Збереження та завантаження побудованої моделі за допомогою модуля pickle.

Тема 6. Нейронні мережі: градієнтні методи оптимізації

Біологічний нейрон, модель МакКаллока-Піттса як лінійний класифікатор. Опції активації. Проблема повноти. Завдання виключної диз'юнкції. Повнота двохшарових мереж у просторі булевих функцій. Алгоритм зворотного розповсюдження помилок. Швидкі методи стохастичного градієнта: Поляка, Нестерова, AdaGrad, RMSProp, AdaDelta, Adam, Nadam, діагональний метод Левенберга-Марквардта. Проблема вибуху градієнта та евристика gradient clipping. Метод випадкових вимкнень нейронів (Dropout). Інтерпретація Dropout. Зворотній Dropout та L2-регуляризація. Функції активації ReLU та PReLU. Проблема «паралічу» мережі. Евристики на формування початкового наближення. Метод пошарового налаштування мережі.

Підбір структури мережі: методи поступового ускладнення мережі, оптимальне проріджування нейронних мереж.

Тема 7. Згорткові нейронні мережі

Обґрунтування глибоких нейронних мереж: виразні можливості, швидкість збіжності при надмірній параметризації. Згорткові нейронні мережі (CNN) для зображень. Згортковий нейрон. Pooling нейрон. Вибірка зображень ImageNet. ResNet: залишкова нейронна мережа (residual NN). Наскрізнi зв'язки між шарами. Згортки для сигналів, текстів, графів, ігор.

Тема 8. Рекурентні нейронні мережі

Рекурентні нейронні мережі (RNN). Навчання рекурентних мереж: Backpropagation Through Time (BPTT). Мережі довготривалої пам'яті (Long short-term memory, LSTM). Рекурентні мережі Gated Recurrent Unit (GRU) та Simple Recurrent Unit (SRU).

Тема 9. Нейронні мережі з навчанням без вчителя

Нейронна мережа Кохонена. Конкурентне навчання, стратегії WTA та WTM. Самоорганізована карта Кохонена. Застосування візуального аналізу даних. Мистецтво інтерпретації карт Кохонена. Автокодувальник. Лінійний AE, SAE, DAE, CAE, RAE, VAE, AE для класифікації, багатошаровий AE. Перед-навчання нейронних мереж. Перенесення навчання. Багатозадачне навчання. Самостійне навчання. Дистиляція моделей або сурогатне моделювання. Навчання з використанням привілейованої інформації (learning using privileged information). Генеративні змагальні мережі (generative adversarial net).

3. Структура навчальної дисципліни

Назви розділів і тем	Кількість годин					
	денна форма					
	усього	у тому числі				
л		п	лаб.	інд.	с. р.	
1	2	3	4	5	6	7
Тема 1. Вступ до алгоритмів для науки даних. Логістична регресія	14	2	2			10
Тема 2. Гребнева та лінійна регресія	15	3	2			10
Тема 3. Мультилінійна регресія	22	6	4			12
Тема 4. Класифікація	22	3	3			16
Тема 5. Кластеризація	17	2	5			10
Тема 6. Нейронні мережі: градієнтні методи оптимізації	14	2	2			10

Тема 7. Згорткові нейронні мережі	24	6	6			12
Тема 8. Рекурентні нейронні мережі	22	4	4			14
Тема 9. Нейронні мережі з навчанням без вчителя	20	4	4			12
Виконання індивідуальної розрахунково-графічної роботи	10					10
Усього годин	180	32	32			116

4. Теми практичних занять

№ з/п	Назва теми	Кількість годин
1	Передбачення достовірності позитивної та негативної події на базі алгоритму логістичної регресії за допомогою пакета <code>sklearn.LogisticRegression</code> . Збереження прогнозованих значень як <code>DataFrame</code> .	2
2	Прогнозування довільних значень за допомогою лінійної регресії. Оцінка якості моделі за допомогою <code>R2 Score</code> .	2
3	Розподіл даних щодо вибірки на тренувальну множину та тестову множину. Застосування крос-валідації. Визначення доцільних факторів для мультилінійної регресії та застосування цього алгоритму для конкретної прикладної задачі.	4
4	Класифікація даних за допомогою алгоритму дерева рішень. Візуалізація побудови рішень на базі впливових факторів за допомогою <code>graphviz</code> . Обґрунтування отриманих результатів за допомогою <code>confusion matrix</code> .	3
5	Застосування алгоритму, що кластеризує дані. Відокремлення зразків в n групах рівної дисперсії. Збереження та загрузка побудованої моделі за допомогою модуля <code>pickle</code> .	2
6	Контрольна робота	3
7	Математичний нейрон Маккаллока - Пітса. Перцептрон Розенблатта. Реалізація градієнтного спуску «з нуля» для простої функції. Порівняння <code>batch</code> , <code>stochastic</code> та <code>mini-batch gradient descent</code> . Використання алгоритмів оптимізації (<code>Momentum</code> , <code>RMSProp</code> , <code>Adam</code>) для навчання простої MLP на датасеті MNIST.	2
8	Реалізація операцій згортки та <code>pooling</code> . Побудова простої CNN для класифікації зображень (MNIST або CIFAR-10). Вплив кількості фільтрів, розмірів ядер, кількості шарів на якість класифікації. Використання <code>Dropout</code> та <code>Batch Normalization</code> у CNN.	6

9	RNN: навчання. RNN: як вирішувати завдання класифікації LSTM. Двонаправлені RNN. Глибокі двонаправлені RNN. Рекурсивні нейронні мережі.	4
10	Використання автоенкодерів для зменшення розмірності даних. Побудова варіаційного автоенкодера (VAE) для генерації зображень. Ознайомлення з генеративно-змагальними мережами (GAN): генерація простих зображень. Кластеризація даних з використанням SOM (Self-Organizing Maps). Порівняння результатів класичних методів зменшення розмірності (PCA, t-SNE) та нейромережевих автоенкодерів.	4
	Разом	32

5. Завдання для самостійної роботи

№ з/п	Види, зміст самостійної роботи	Кількість годин
1	Виконання домашніх завдань з визначенням цільових напрямків моделі даних	10
2	Виконання домашніх завдань з розробки моделі даних, яка задовольняє вимогам масштабування	10
3	Виконання домашніх завдань з побудови структури алгоритму на базі моделі даних	12
4	Виконання домашніх завдань з реалізації та аналізу алгоритмів на Python	16
5	Виконання домашніх завдань по збереженню та завантаженню побудованої моделі за допомогою модуля pickle.	10
6	Виконання домашніх завдань з порівняння швидкості збіжності різних оптимізаторів (SGD, Momentum, Adam), реалізації власного оптимізатора.	10
7	Виконання домашніх завдань зі згорткових нейронних мереж	12
8	Виконання домашніх завдань з рекурентних нейронних мереж	14
9	Виконання домашніх завдань на використання автоенкодерів у різних задачах	12
10	Виконання індивідуальної розрахунково-графічної роботи	10
	Разом	116

6. Методи навчання

Частково-пошуковий, дослідницький, пояснювально-ілюстративний, репродуктивний. Лекційні та практичні заняття.

7. Методи контролю

- перевірка виконання домашніх завдань
- поточне опитування за лекційним матеріалом
- перевірка індивідуального завдання
- перевірка контрольної роботи
- перевірка звітів-результатів самостійної роботи студентів
- залік

8. Схема нарахування балів

Поточний контроль, самостійна робота, індивідуальні завдання				Залік	Сума
T1-T9	Розрахунково-графічна робота	Контрольна робота	Разом		
20	20	20	60	40	100

Критерії оцінювання

Поточний контроль:

бали нараховуються протягом курсу за виконання домашніх завдань, контрольної роботи і активність студентів під час лекцій (через їх поточне опитування) та практичних занять (де в тому числі схвалюється, враховується та винагороджується балами їх самостійна робота й ініціативи).

За контрольну роботу бали нараховуються таким чином:

- максимальний бал у разі правильної обґрунтованої відповіді без помилок;
- за незначні помилки оцінка зменшується від 10 до 30 відсотків;
- за значні логічні помилки оцінка зменшується до 50 відсотків, якщо хід розв'язання/міркування в цілому правильний,
- у разі частково правильних міркувань за відсутності обґрунтованої відповіді виставляється до 30 відсотків від максимальної кількості балів
- якщо відповідь не відповідає жодному з критеріїв, які сформульовані вище, – виставляється 0 балів.

У межах розрахунково-графічного завдання виконується комплексний аналіз даних із застосуванням сучасних методів машинного навчання. Робота передбачає побудову та оцінювання моделей регресії, класифікації та кластеризації, включаючи логістичну регресію, лінійну й мультилінійну регресію, дерево рішень і алгоритми кластеризації. Особлива увага приділяється

підготовці даних, розподілу вибірки на тренувальну та тестову, використанню крос-валідації, оцінці якості моделей і візуалізації результатів.

Для допуску до складання підсумкового контролю (заліку) здобувач вищої освіти повинен набрати не менше 10 балів з навчальної дисципліни протягом семестру.

Залік:

залікова робота складається із двох частин/блоків, що відповідає двом питанням/завданням. Максимальна оцінка за кожне – 20 балів, разом за обидва – 40 балів.

Для отримання максимальної оцінки (20 балів) за одне питання потрібно вміти сформулювати відповідні означення та/або навести алгоритм(и), надати аналіз щодо структури даних, на яких буде реалізовуватися цей алгоритм та обов'язково продемонструвати володіння синтаксисом відповідного алгоритму на прикладі(ах). Замість прикладів можуть бути наведені міркування щодо змісту відповідних понять і результатів. Викладення основних концепцій та принципів роботи базового(их) алгоритму(ів) науки даних та логіки (побудова діаграм та альтернативна візуалізація результатів) є обов'язковою.

Формулювання неповне або зі значними помилками: 5-15 балів.

Формулювання без прикладів або міркувань: 10-15 балів.

Шкала оцінювання (дворінева)

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка
50-100	зараховано
1-49	не зараховано

10. Рекомендована література

Основна література

1. Субботін С. О. Нейронні мережі : теорія та практика: навч. посіб. – Житомир : Вид. О. О. Євенок, 2020. – 184 с.
2. Кононова К. Ю. Машинне навчання. Методи та моделі. Підручник. – Харків: ХНУ ім. В. Н. Каразіна, 2020. – 279 с.
3. Терейковський І. А., Бушуєв Д. А., Терейковська Л. О. Штучні нейронні мережі: Базові положення. Навчальний посібник. Київ: КПІ ім. Ігоря Сікорського, 2022. – 123 с.
4. Richert W., Coelho L. P. Building Machine Learning Systems with Python. – Packt, 2013. – 290 p.
5. Chollet F. Deep Learning with Python 1st Edition. – Manning, 2017. — 384 p.

Допоміжна література

1. Троцько В.В. Методи штучного інтелекту: навчально-методичний і практичний посібник. – Київ: Університет економіки та права «КРОК», 2020. – 86 с.
2. Rojas R. Neural Networks. A systematic introduction. – Springer-Verlag, 1996.
3. Steiner C. Automate This: How Algorithms Came to Rule Our World. – Portfolio Hardcover, 2012. – 256 p.
4. Haykin S. Neural Networks and Learning Machines. – Prentice Hall, 2009. – 906 p.

Посилання на інформаційні ресурси в Інтернеті, відео-лекції, інше мет. заб.

1. Ресурс з різними датасетами, що можна використовувати при навчанні: <https://www.kaggle.com/>
2. Web app that allows users to test the artificial intelligence (AI) algorithm with TensorFlow machine learning library: <https://playground.tensorflow.org/>