МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ОДЕСЬКА ПОЛІТЕХНІКА»
КАФЕДРА ЕКОЛОГІЧНОЇ БЕЗПЕКИ ТА ГІДРАВЛІКИ

# Конспект лекцій
# з дисципліни ” Методи обробки геопросторових даних”

для здобувачів першого (бакалаврського) рівня вищої освіти
за спеціальністю 101 Екологія
освітня програма «Екологічна безпека»

Одеса 2025

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ОДЕСЬКА ПОЛІТЕХНІКА»
КАФЕДРА ЕКОЛОГІЧНОЇ БЕЗПЕКИ ТА ГІДРАВЛІКИ

# Конспект лекцій
# з дисципліни ” Методи обробки геопросторових даних”

для здобувачів першого (бакалаврського) рівня вищої освіти
за спеціальністю 101 Екологія
освітня програма «Екологічна безпека»

Затверджено
на засіданні кафедри
екологічної безпеки та
гідравліки
Протокол № 5 від
29.12.2025 р.

Одеса 2025

Конспект лекцій з дисципліни ” Методи обробки геопросторових даних” для студентів зі спезіальності 101 - Екологія./ Укл.: С.Ю. Смик, С.Г. Антощук, Д.В. Кошутіна. Одеса: ОНПУ, 2025. – 99 с.

Укладачі: Смик С.Ю., доцент
Антощук С.Г., професор
Кошутіна Д.В., асистент

У сучасному інформаційному суспільстві методи обробки геопросторових даних перетворилися з допоміжних інструментів на фундаментальну основу стратегічного управління, наукових досліджень та забезпечення національної безпеки. Експоненціальне зростання обсягів інформації, що генерується супутниковими угрупованнями, мережами датчиків Інтернету речей (IoT), безпілотними літальними апаратами та системами дистанційного зондування, створює нову цифрову реальність. Для фахівця в галузі екології та комп'ютерних наук глибоке розуміння алгоритмів інтелектуального аналізу геоданих є критичною компетенцією, що дозволяє конвертувати терабайти «сирих» сигналів у дієві стратегії порятунку біосфери.

Актуальність та фундаментальна важливість вивчення даної дисципліни зумовлена наступними стратегічними чинниками:

- необхідністю оперативного реагування на глобальні кліматичні виклики, де предиктивні моделі дозволяють не лише фіксувати зміни температури чи рівня світового океану, а й прогнозувати каскадні наслідки для локальних екосистем на десятиліття вперед;
- стрімким розвитком концепції «цифрових двійників» (Digital Twins) територій та природних об'єктів, що створює віртуальні полігони для моделювання антропогенного навантаження, дозволяючи проводити стрес-тестування екосистем без ризику незворотного фізичного пошкодження середовища;
- інтеграцією передових архітектур штучного інтелекту в системи підтримки прийняття рішень, що суттєво мінімізує вплив когнітивних упереджень та людського фактора при проведенні екологічної експертизи та оцінці ризиків на промислових об'єктах;
- потребою у високотехнологічній обробці великих масивів даних (Big Environmental Data), де класичні методи статистики поступаються місцем алгоритмам глибокого навчання, здатним виявляти

латентні кореляції між забрудненням атмосфери, станом гідросфери та деградацією біорізноманіття;

- критичним значенням точних просторових прогнозів для сталого розвитку аграрного сектора, де технології точного землеробства дозволяють оптимізувати використання ресурсів та мінімізувати хімічне навантаження на ґрунти;
- необхідністю побудови прозорих та верифікованих систем екологічного моніторингу, які забезпечують суспільство та державні органи достовірною інформацією про стан довкілля, унеможливлюючи маніпуляції та приховування фактів екологічних правопорушень;
- формуванням нової етики управління планетарними ресурсами, яка базується на математично обґрунтованому балансі між економічним прогресом та збереженням життєздатності природного капіталу.

Таким чином, дана дисципліна виступає інтелектуальним містком між теоретичною екологією та практичною інженерією даних.

# ЛЕКЦІЯ №1: ВСТУП ДО МЕТОДІВ ОБРОБКИ ПРОСТОРОВИХ ДАНИХ В ЕКОЛОГІЧНОМУ МОНІТОРИНГУ ТА МОДЕЛЮВАННІ

## 1. Фундаментальне значення обробки геопросторових даних для екології

Сучасна екологія пройшла шлях трансформації від дескриптивної (описової) науки до високотехнологічної галузі, що базується на об’єктивних інструментальних вимірюваннях. У контексті глобальних кліматичних змін та антропогенного тиску на біосферу, традиційних методів спостереження стає недостатньо. Саме тому методи обробки геопросторових даних стають ключовим інструментом для фахівця-еколога, що зумовлено наступними чинниками:

- необхідністю аналізу процесів, що мають просторову прив’язку, оскільки будь-яка екологічна подія (забруднення акваторії, деградація ґрунтів, міграція популяцій) відбувається у конкретних географічних координатах;
- можливістю інтеграції різнорідної інформації, що дозволяє поєднувати супутникові знімки, дані наземних метеостанцій, результати хімічних аналізів проб та цифрові моделі рельєфу в єдину аналітичну систему;
- потребою у масштабуванні досліджень, що забезпечує перехід від локального моніторингу окремого підприємства до оцінки стану екосистем на рівні регіонів, країн та континентів;
- автоматизацією виявлення критичних змін у довкіллі, таких як незаконні вирубки лісів, розширення площ звалищ або динаміка танення льодовиків, що фізично неможливо охопити лише зусиллями польових інспекторів;
- переходом до прогностичного моделювання, що дає змогу не лише констатувати факт екологічного лиха, а й прогнозувати поширення

забруднюючих речовин в атмосфері чи гідросфері на основі математичних моделей.

Таким чином, володіння методами обробки геоданих дозволяє екологу перетворити масиви «сирих» цифр на дієвий інструмент для прийняття управлінських рішень та розробки стратегій сталого розвитку.

**2. Мета та задачі навчальної дисципліни**

**Мета курсу** полягає у формуванні у студентів цілісної системи теоретичних знань та практичних навичок щодо застосування математичних алгоритмів обробки просторової інформації для вирішення прикладних задач екологічної безпеки та раціонального природокористування.

До основних задач дисципліни належать:

- вивчення специфіки архітектури геопросторових даних та методів їх консолідації з різних джерел (ДЗЗ, ГІС-мережі, сенсорні системи);
- засвоєння принципів збору, верифікації та очищення вихідних даних від шумів та інструментальних похибок;
- опанування алгоритмів класифікації природних об’єктів та детектування змін ландшафтів;
- вивчення методів аналізу часових рядів для моніторингу динаміки екологічних процесів;
- формування навичок візуалізації результатів аналізу у вигляді тематичних карт та інтерактивних моделей.

**3. Концепція Big Environmental Data та «4 V» в екологічному контексті**

Стрімкий розвиток автоматизованих систем збору інформації призвів до виникнення феномену «великих екологічних даних». Обробка такої інформації вимагає спеціальних підходів, що описуються через чотири фундаментальні характеристики:

**Volume (Об'єм):** щоденне надходження терабайтів супутникових даних високої роздільної здатності та мільйонів записів з автоматичних постів контролю якості повітря;

**Variety (Різнорідність):** поєднання структурованих числових даних (концентрації речовин), растрових зображень (спектральні знімки), векторних шарів (межі заповідників) та неструктурованих звітів;

**Velocity (Швидкість):** необхідність обробки потоків даних у режимі реального часу для миттєвого реагування на техногенні аварії;

**Veracity (Достовірність):** складність фільтрації похибок вимірювальних приладів, що працюють у несприятливих польових умовах.

**4. Попередня обробка та математичне перетворення екологічних показників**

На практиці еколог отримує дані у різних одиницях вимірювання ( мг/м$^3$, мг/л, градуси Цельсія, бали бонітету), що унеможливлює їх пряме порівняння. Для коректної роботи алгоритмів необхідно виконати попередню обробку.

**4.1. Масштабування**

Це процес приведення ознак до певного діапазону (зазвичай $[0, 1]$). Найбільш поширеним є мінімаксне шкалювання:

$$x_i^{'} = \frac{x_i - (x)}{(x) - (x)}$$

*Приклад:* приведення значень індексу вегетації (NDVI) та рівня вологості ґрунту до єдиної шкали для аналізу їх кореляції.

**4.2. Бінаризація**

Процес перетворення числових значень у логічні (0 або 1). В екології це найчастіше використовується для фіксації перевищення ГДК (гранично допустимої концентрації):

- якщо $C >$ ГДК, то результат $-$ 1 (небезпека);
- якщо $C \leq$ ГДК, то результат $-$ 0 (норма).

**4.3. Виняток середнього**

Методика, що дозволяє центрувати ознаки на нулі. Це критично важливо для усунення систематичних зміщень у показах сенсорів, що тривалий час працюють без калібрування.

**4.4. Нормалізація**

Зміна значень у векторі ознак таким чином, щоб для їх вимірювання можна було використовувати евклідову норму ($L2$-норма):

$$|x|_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$$

Цей метод дозволяє порівнювати «екологічні профілі» різних територій незалежно від абсолютної величини вимірюваних параметрів.

**5. Сфери прикладного застосування методів обробки геоданих**

- **Моніторинг лісових ресурсів:** оцінка біомаси, виявлення вогнищ захворювань рослин та прогнозування пожежної небезпеки;
- **Охорона водних екосистем:** моделювання розповсюдження плям нафтопродуктів та аналіз евтрофікації (цвітіння) водойм;
- **Агроекологія:** точне землеробство, оцінка деградації (ерозії) ґрунтів та оптимізація внесення добрив;
- **Урбоекологія:** картографування теплових островів міст та аналіз поширення акустичного забруднення.

**Контрольні питання для самоперевірки:**

1. Поясніть необхідність масштабування даних при одночасному аналізі концентрації пилу та швидкості вітру.
2. Які переваги дає використання супутникових даних порівняно з традиційними методами хімічного аналізу проб?
3. Опишіть характеристики «Big Data» на прикладі системи моніторингу якості води у річці Дніпро.

# ЛЕКЦІЯ №2: ПОСТАНОВКА ЗАДАЧІ РОЗПІЗНАВАННЯ ТА КЛАСИФІКАЦІЇ В ЕКОЛОГІЧНОМУ МОНІТОРИНГУ

## 1. Сутність розпізнавання у системах обробки геоданих

Процес розпізнавання становить фундаментальну основу функціонування будь-якої інтелектуальної системи екологічного моніторингу, що базується на обробці просторово-координованої інформації. У найбільш загальному розумінні, розпізнавання – це процедура ідентифікації поточного стану або функціонального типу природного чи антропогенного об’єкта на підставі аналізу його цифрової сигнатури. В екологічному контексті це означає формалізоване віднесення спостережуваних параметрів навколишнього середовища (спектральних сигнатур, концентрацій полютантів, біоакустичних сигналів) до заздалегідь визначених категорій або екологічних станів.

Ключовими прикладними задачами розпізнавання у сучасній екології є:

- дешифрування мультиспектральних та гіперспектральних даних дистанційного зондування Землі (ДЗЗ) для прецизійної диференціації типів підстильної поверхні, включаючи хвойні та листяні лісові масиви, водна дзеркала з різним ступенем евтрофікації, урбанізовану забудову та сільськогосподарські угіддя;
- автоматизована діагностика фітопатологічного стану лісів, що дозволяє виявляти зони пошкодження ентомологічними шкідниками, осередки хлорозу та дефоліації або латентні ознаки деградації внаслідок накопичення токсикантів у ґрунтах;
- ідентифікація біологічних об’єктів за специфічними акустичними чи візуальними дескрипторами, включаючи розпізнавання видів за вокалізацією (біоакустичний моніторинг) або автоматичну

каталогізацію особин на кадрах фотопасток за допомогою методів глибокого навчання;

− динамічне районування територій за рівнем інтегрального антропогенного навантаження та побудова карт екологічного ризику на основі багатофакторного аналізу;

− сегментація супутникових знімків у реальному часі для визначення динаміки меж розливу нафтопродуктів, моніторингу просування фронту лісових пожеж або прогнозування зон затоплення під час паводків.

Методологічно процес трансформації сирих даних у знання для підтримки екологічних рішень є багаторівневим і базується на ієрархії DIKW (Data, Information, Knowledge, Wisdom). Він починається з первинної фіксації фізичних сигналів датчиками, які в подальшому структурується в інформаційні одиниці з обов'язковою прив’язкою до часових та просторових координат. На заключному етапі, через механізми логічного висновку та розпізнавання закономірностей, формуються знання про системну поведінку екосистеми та прогнозуються сценарії її розвитку.

**2. Категоріальний апарат: Образ, Ознака та Клас**

Побудова строгих математичних моделей екологічних систем вимагає використання специфічного понятійного апарату, що описує взаємодію об’єктів у просторі станів:

− **образ −** це теоретична абстракція, що об’єднує сукупність об’єктів, які володіють ідентичним або максимально близьким набором фундаментальних властивостей (еталонів) з точки зору конкретної екологічної задачі;

− **реалізація образу −** конкретний фізичний об’єкт, точка інструментального заміру або елементарна ділянка ландшафту (піксель знімка), що фактично належить до певної категорії та має унікальний набір параметрів у момент фіксації;

– **ознака (дескриптор)** – кількісна або якісна міра властивості об’єкта, яка має бути дискримінативною, тобто дозволяти математично відокремити один клас від іншого з мінімальною імовірністю похибки першого та другого роду;

– **простір ознак** – абстрактний $n$-вимірний метричний простір, де кожному об’єкту відповідає вектор (точка), координати якого є значеннями його ознак. Розмірність цього простору визначає складність аналітичної моделі.

Важливо розуміти діалектичну єдність «образу» та «ознаки»: завдання образу автоматично породжує потребу у специфічних ознаках для його ідентифікації (наприклад, для образу «здоровий ліс» критичною ознакою є високий рівень відбиття у ближньому інфрачервоному діапазоні), і навпаки – набір доступних ознак обмежує алфавіт образів, які система здатна розрізнити.

**3. Розгорнута типологія ознак екологічних об’єктів**

Ефективність систем розпізнавання безпосередньо залежить від коректного вибору типу ознак, що враховують фізичну та хімічну природу екологічних процесів:

– детерміновані ознаки – параметри з фіксованими значеннями, де інструментальна похибка є нехтовно малою відносно масштабу задачі, наприклад, площа природно-заповідного фонду або географічна висота джерела викидів над рівнем моря;

– імовірнісні ознаки – характеристики, що мають стохастичну природу і описуються через статистичні розподіли, такі як добовий хід концентрації приземного озону або флуктуації щільності популяції у відповідь на метеорологічні зміни;

– логічні (бінарні або порядкові) ознаки – дискретні стани, що вказують на наявність або відсутність критичної умови, наприклад, статус

території як об’єкта Смарагдової мережі або ступінь стійкості ландшафту до ерозії (низька, середня, висока);

− структурно-геометричні ознаки − опис просторової конфігурації, топологічних зв’язків та фрактальної розмірності об'єктів, що є критично важливим для аналізу фрагментації екосистем або розгалуженості гідрографічної мережі.

**4. Гіпотеза компактності та метричний аналіз**

Математична можливість автоматичної сепарації екологічних станів базується на фундаментальній **гіпотезі компактності**. Ця гіпотеза постулює: якщо ми обрали релевантний та інформативний набір ознак, то об’єкти одного природного класу будуть розташовані у просторі ознак максимально щільно, утворюючи обмежені та компактні згустки (кластери).

Для забезпечення коректної класифікації мають виконуватися наступні умови:

− внутрішньокласова відстань (відстань між реалізаціями одного образу) повинна бути суттєво меншою за міжкласову;

− міжкласова відстань (відстань між центрами груп різних образів) має бути максимальною, що дозволяє провести між ними стійку роздільну гіперповерхню;

− класи не повинні мати критичних областей перекриття. У разі виникнення «сірих зон» необхідно застосовувати методи зниження розмірності або переходити у простір ознак вищого порядку.

Близькість об’єктів у екологічному моделюванні оцінюється за допомогою метрик, вибір яких залежить від природи даних:

− евклідова відстань − класична міра, що визначає пряму лінію у просторі ознак. Найбільш ефективна при роботі з нормалізованими даними однакової фізичної природи;

− манхеттенська відстань (метрика таксі) − сума абсолютних різниць координат. Часто є більш стійкою до поодиноких аномальних викидів у даних (outliers), що характерно для польових екологічних датчиків;

− відстань Чебишова − визначає максимальну різницю між значеннями ознак об’єктів. Використовується, коли критичним є найгірший показник серед усіх параметрів моніторингу;

− відстань Махаланобіса − враховує кореляційні зв’язки між ознаками та масштаби їх варіації, що дозволяє ефективно виявляти аномалії в хімічному складі середовища, де вміст одного металу корелює з іншим.

**5. Базові алгоритмічні підходи до класифікації**

Прийняття рішення про належність об’єкта до конкретної екологічної категорії здійснюється за допомогою вирішальних правил (decision rules).

**5.1. Класифікація на основі еталонних центрів (барицентрів)**

Цей метод передбачає попереднє обчислення «ідеального образу» класу шляхом знаходження середнього арифметичного вектора всіх ознак еталонних об’єктів (centroid). Система порівнює кожен новий вектор вимірювань із центрами доступних класів і відносить його до того, відстань до якого є мінімальною. *Переваги:* висока швидкість обчислень та стійкість до випадкових шумів у вибірці. *Застосування:* оперативне районування великих територій за типом антропогенного навантаження на основі інтегральних екологічних індексів.

**5.2. Алгоритм k-найближчих сусідів (k-Nearest Neighbors, k-NN)**

Об’єкт класифікується не на основі абстрактного геометричного центру, а за подібністю до конкретних реальних реалізацій, що вже зафіксовані в базі даних. Об’єкт відноситься до того класу, який є найбільш масовим серед його $k$ найближчих сусідів у просторі ознак.

*Недоліки:* значні обчислювальні витрати при великих обсягах даних та чутливість до нерелевантних ознак. *Застосування:* прецизійна класифікація біотопів та ідентифікація патологій рослинності за складними спектральними профілями.

**6. Геометрична інтерпретація та роздільні поверхні**

В основі автоматизованого екологічного районування лежить концепція роздільної поверхні (decision boundary). Математично це функція $f(X)$, яка розділяє простір ознак на підобласті, кожна з яких відповідає певному екологічному стану.

Для лінійного класифікатора вирішальне правило має вигляд:

$$D(X) = \sum_{i=1}^{n} w_i x_i + w_0$$

де $w_i$ – вагові коефіцієнти, що визначають значущість кожної ознаки для ідентифікації класу. Знак функції $D(X)$ визначає приналежність об'єкта до тієї чи іншої сторони роздільної гіперповерхні.

Диференціація зон ґрунтового забруднення за концентрацією важких металів ($x_1$) та пестицидів ($x_2$). Побудова роздільної лінії між класом «чисті ґрунти» та «забруднені ґрунти» дозволяє автоматично генерувати карти придатності земель для аграрного використання. Будь-яка нова проба $O(x_1, x_2)$ негайно отримує статус безпеки без залучення експерта, що критично важливо для оперативного моніторингу великих агрохолдингів або природно-заповідних фондів.

**Контрольні питання для закріплення матеріалу:**

1. обґрунтуйте вибір метрики Махаланобіса при аналізі забруднення підземних вод, де концентрації різних іонів мають високий ступінь кореляції;

2. яким чином «прокляття розмірності» (curse of dimensionality) впливає на виконання гіпотези компактності в екологічних моделях з великою кількістю дескрипторів;

3. опишіть алгоритм вибору вагових коефіцієнтів $w_i$ для лінійного класифікатора при оцінці пожежної небезпеки лісів;

4. поясніть фізичний зміст «структурної ознаки» на прикладі аналізу розгалуженості річкової мережі та її впливу на швидкість розповсюдження рідких забруднювачів.

# ЛЕКЦІЯ №3: ОПТИМІЗАЦІЯ, РЕГРЕСІЙНЕ МОДЕЛЮВАННЯ ТА ГРАДІЄНТНІ МЕТОДИ В ЕКОЛОГІЧНИХ ДОСЛІДЖЕННЯХ

## 1. Фундаментальна роль математичної оптимізації в геоекології

Екологічні процеси за своєю природою є багатофакторними та стохастичними. Для фахівця, що здійснює обробку геопросторових даних, критично важливо перейти від простого накопичення результатів спостережень до побудови предиктивних моделей. Оптимізація у цьому контексті виступає як механізм «калібрування» моделі, що дозволяє знайти найкращі параметри системи за умови мінімальної розбіжності з реальністю.

Історично методи регресії та оптимізації почали свій розвиток з робіт Карла Фрідріха Гаусса (1821 рік), який запропонував метод найменших квадратів, та Френсіса Гальтона (1889 рік), який ввів сам термін «регресія» під час вивчення спадковості. Сьогодні ці методи є основою для:

- моделювання шлейфів розсіювання забруднюючих речовин в атмосферному повітрі залежно від геоморфологічних умов та рози вітрів;
- аналізу динаміки «теплових островів» у міському середовищі на основі даних супутникового термального моніторингу;
- розрахунку екологічної ємності територій та прогнозування деградації біорізноманіття;
- оптимізації розміщення мереж автоматизованих постів екологічного контролю для максимізації охоплення території при обмежених ресурсах.

## 2. Математична постановка задач оптимізації в екологічному моделюванні

Процес оптимізації в обробці геоданих − це процедура налаштування внутрішніх параметрів моделі (ваг) для досягнення екстремуму (мінімуму або максимуму) цільової функції.

### 2.1. Цільова функція та функція втрат

В екологічних дослідженнях ми найчастіше мінімізуємо функцію втрат ($Loss\ Function$), яка кількісно оцінює помилку моделі. Основними критеріями якості оптимізації є:

− адекватність − здатність моделі відображати реальні екологічні закономірності;

− стійкість (робастність) − нечутливість моделі до поодиноких викидів, спричинених збоями в роботі польових датчиків;

− збіжність − гарантоване знаходження глобального мінімуму помилки за обмежену кількість ітерацій обчислення.

### 2.2. Проблема локальних мінімумів

При обробці складних геопросторових масивів (наприклад, цифрових моделей рельєфу або карт ґрунтів) функція втрат може мати багато «пасток» − локальних мінімумів. Завданням методів оптимізації є знаходження саме глобального оптимуму, що забезпечує найбільш точну інтерпретацію стану природної системи.

## 3. Регресійний аналіз як метод апроксимації геоданих

Регресія в геоекології − це статистичний метод моделювання залежності між незалежними змінними (факторами середовища) та залежною змінною (екологічним індикатором).

### 3.1. Лінійна регресія: принципи та припущення

Лінійна модель передбачає, що екологічний відгук змінюється пропорційно впливу факторів. Математична форма:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n +$$

де:

– $y$ – цільовий екологічний показник (наприклад, концентрація пилу $PM2.5$);

– $x_i$ – предиктори (інтенсивність трафіку, вологість повітря, висота над рівнем моря);

– $\beta_i$ – коефіцієнти, що визначають силу та напрямок впливу кожного фактора;

– $\epsilon$ – випадкова похибка (шум), що враховує невраховані мікрокліматичні варіації.

### 3.2. Логістична регресія та імовірнісний аналіз ризиків

Коли завдання полягає у визначенні імовірності настання критичної події, використовується логістична регресія. Результат моделювання знаходиться в діапазоні $[0, 1]$. *Сфери застосування:*

– оцінка імовірності самозаймання торфовищ;

– прогнозування поширення інвазивних видів рослин;

– визначення зон ризику затоплення прибережних територій.

Центральним елементом моделі є сигмоїдна функція:

$$\sigma(z) = \frac{1}{1+e^{-\left(\beta_0+\sum\beta_i x_i\right)}}$$

### 3.3. Регресія на основі опорних векторів (SVR)

SVR є потужним інструментом для обробки нелінійних та зашумлених геоданих. На відміну від класичної регресії, SVR намагається знайти таку функцію, яка б не просто мінімізувала помилку, а утримувала більшість значень вимірювань у межах певної «трубки» відхилення $\epsilon$. Це дозволяє ігнорувати незначні коливання приладів, фокусуючись на загальному тренді екологічного процесу.

## 4. Метод найменших квадратів (МНК) та аналітичне рішення

МНК є базовим алгоритмом для знаходження коефіцієнтів лінійної регресії. Його суть полягає у мінімізації суми квадратів відхилень прогнозованих значень від реальних даних моніторингу:

$$E(\beta) = \sum_{i=1}^{n} \left(y_i - \left(\beta_0 + \beta_1 x_i\right)\right)^2 \rightarrow min$$екологічних задачах МНК дозволяє:

- будувати лінії трендів багаторічних спостережень за рівнем озонового шару;
- встановлювати кореляції між дозами внесення добрив та залишковою концентрацією нітратів у підземних водах;
- усереднювати дані з різних метеостанцій для побудови узагальненої карти опадів.

**5. Градієнтний спуск: алгоритмічна основа навчання моделей**

Для обробки сучасних «великих екологічних даних» (терабайти супутникових знімків), аналітичне розв'язання МНК стає обчислювально неможливим. У таких випадках застосовується ітераційний метод градієнтного спуску.

**5.1. Механіка процесу**

Градієнтний спуск − це покрокове наближення до мінімуму функції втрат шляхом руху в напрямку, протилежному вектору градієнта. В кожній точці обчислюється похідна, яка вказує, як саме потрібно змінити параметри моделі (β), щоб помилка зменшилася.

Формула оновлення параметрів:

$$\beta_{new} = \beta_{old} - \alpha \cdot \frac{\partial J(\beta)}{\partial \beta}$$

де:

- α − швидкість навчання (Learning Rate), що визначає розмір кроку;

– $\frac{\partial J(\beta)}{\partial \beta}$ – часткова похідна функції втрат, що вказує напрямок найшвидшого зростання помилки.

**5.2. Роль швидкості навчання (α)**

Вибір α є критичним для еколога-аналітика:

– якщо α занадто велике, алгоритм може «перестрибнути» мінімум, що призведе до розбіжності моделі;

– якщо α занадто мале, процес обробки геоданих триватиме надто довго, що недопустимо при моніторингу надзвичайних ситуацій.

**5.3. Різновиди градієнтних методів**

– пакетний (Batch) – використовує всі точки вимірювань для кожного кроку, забезпечує стабільність, але вимагає багато пам'яті;

– стохастичний (SGD) – оновлює ваги після кожної окремої точки даних (наприклад, після кожного заміру сенсора), що забезпечує швидкість, але результат може сильно вагатися;

– міні-пакетний (Mini-batch) – компромісний варіант, де дані обробляються групами, що є найбільш ефективним для систем реального часу.

**6. Критерії якості та функції втрат у геоекологічних задачах**

Вибір метрики оцінки помилки залежить від специфіки екологічного об'єкта:

– середньоквадратична помилка ($MSE$) – підсилює вплив великих відхилень через піднесення до квадрату. Рекомендується для задач, де пропуск критичного забруднення є катастрофічним;

– середня абсолютна помилка ($MAE$) – менш чутлива до аномальних викидів. Використовується при первинній обробці сирих даних з недосконалих польових датчиків;

– крос-ентропія ($Log\ Loss$) – стандарт для моделей логістичної регресії, що оцінюють ризики екологічних катастроф, де важливо враховувати впевненість моделі у прогнозі.

**Контрольні питання для самостійного опрацювання:**

1. обґрунтуйте доцільність використання регуляризації в регресійних моделях при аналізі стану лісів за допомогою супутникових даних;
2. поясніть, як за допомогою градієнтного спуску можна автоматизувати пошук оптимальних коефіцієнтів у моделі розповсюдження радіонуклідів;
3. порівняйте ефективність логістичної регресії та методу опорних векторів для задачі класифікації ступеня деградації ґрунтів;
4. які екологічні фактори можуть викликати появу «викидів» у даних моніторингу та як це впливає на вибір між метриками $MSE$ та $MAE$;
5. опишіть фізичний зміст параметрів $\beta_0$ та $\beta_1$ у моделі залежності температури повітря від географічної широти.

# ЛЕКЦІЯ №4: КЛАСТЕРИЗАЦІЯ ГЕОПРОСТОРОВИХ ДАНИХ ТА АЛГОРИТМ K-СЕРЕДНІХ В ЕКОЛОГІЧНИХ ЗАДАЧАХ

1. Фундаментальні відмінності між кластеризацією та класифікацією

В аналізі геопросторових даних часто виникає змішування понять «класифікація» та «кластеризація». Проте для побудови коректних екологічних моделей необхідно чітко розрізняти ці підходи:

- класифікація – це процес віднесення об'єкта до одного із заздалегідь відомих класів. Наприклад, маючи навчальну вибірку пікселів, де точно відомо, що це «сосновий ліс» або «заболочена територія», алгоритм вчиться розпізнавати ці образи на нових знімках;
- кластеризація – це процес виявлення природних угруповань в даних без апріорних знань про їх природу. Алгоритм групує об'єкти виключно на основі їх математичної близькості у просторі ознак.

В екології кластеризація є первинним етапом дослідження, коли фахівець не знає структури забруднення території або меж біотопів і хоче, щоб система самостійно підказала зони неоднорідності.

2. Розгорнута типологія методів кластеризації

Для обробки геопросторової інформації використовуються різні математичні стратегії групування, кожна з яких має свою специфіку застосування.

2.1. Розділяючі методи

Найпоширенішим представником є алгоритм k-середніх (K-means). Дані розбиваються на $k$ непересічних груп.

- переваги: висока швидкість обробки великих растрових масивів ДЗЗ;
- обмеження: алгоритм вимагає заздалегідь вказати кількість кластерів і погано працює з групами складної (не сферичної) форми.

2.2. Ієрархічні методи

Будують послідовність вкладених розбиттів, що візуалізується у вигляді дендрограми.

− агломеративні (знизу вгору) − кожен об'єкт починає як окремий кластер, а потім вони об'єднуються;

− дивізивні (зверху вниз) − вся вибірка вважається одним кластером, який поступово розщеплюється.

− застосування: вивчення філогенетичних зв'язків видів або ієрархічне структурування річкових басейнів за ступенем антропогенного навантаження.

2.3. Методи на основі щільності

Ключовим алгоритмом є DBSCAN. Кластер визначається як область простору, де щільність точок (об'єктів) перевищує певний поріг.

− переваги: здатність знаходити кластери будь-якої геометричної форми (наприклад, звивистий шлейф забруднення вздовж річки) та автоматичне відсікання шумів (аномалій);

− недоліки: чутливість до різної щільності даних у різних частинах вибірки.

2.4. Грід-методи

Простір ознак розбивається на скінченну кількість осередків (сітку). Обробка ведеться не на рівні точок, а на рівні осередків.

− застосування: швидкий аналіз глобальних кліматичних даних та побудова теплових карт великих міст.

3. Чинники, що впливають на точність кластеризації геоданих

Точність та інтерпретованість результатів в екологічному моніторингу критично залежать від підготовки даних та вибору параметрів моделі.

3.1. Вибір метрики відстані

Форма кластера та його склад змінюються залежно від того, як ми вимірюємо близькість:

− евклідова відстань − ідеальна для фізичних координат;

− відстань Махаланобіса − незамінна при аналізі хімічного складу ґрунтів, оскільки враховує кореляцію між елементами (наприклад, присутність кадмію часто супроводжує цинк);

− косинусна подібність − використовується для порівняння спектральних профілів рослинності.

3.2. Масштабування та нормалізація ознак

Екологічні дані часто мають різні порядки величин: концентрація металів може вимірюватися у міліграмах, а площа лісу − у гектарах. Без попереднього Z-масштабування ознака з найбільшим числовим значенням повністю поглине вплив інших факторів, що призведе до хибних результатів.

3.3. Вплив шумів та аномальних викидів

У геопросторовій обробці викиди можуть бути як помилками датчиків, так і ознаками локальних екологічних катастроф. Алгоритми типу K-means вкрай нестійкі до викидів: один аномальний замір може змістити центр кластера на кілометри, спотворюючи всю картину районування.

3.4. Прокляття розмірності

При збільшенні кількості параметрів (наприклад, аналіз 100+ спектральних каналів гіперспектральної зйомки) відстань між будь-якими двома точками стає майже однаковою. У таких умовах кластеризація втрачає сенс, що вимагає попереднього застосування PCA для зниження розмірності.

4. Математичне обґрунтування алгоритму k-середніх (K-means)

Алгоритм реалізує ітераційний процес мінімізації сумарного квадратичного відхилення точок від центрів кластерів.

4.1. Формалізація кроків

1. Ініціалізація центроїдів $\{\mu_1, ..., \mu_k\}$;

2. Призначення об'єктів: $S_j^{(t)} = \{x_p : |x_p - \mu_j^{(t)}|^2 \leq |x_p - \mu_i^{(t)}|^2 \text{ для всіх } i = 1, ..., k\}$;

3. Оновлення центроїдів: $\mu_j^{(t+1)} = \frac{1}{|S_j^{(t)}|} \sum_{x_i \in S_j^{(t)}} x_i$;

4. Зупинка при стабілізації центроїдів.

4.2. Критерії якості розбиття

− сума квадратів внутрішньокластерних відстаней (WCSS) − чим менше значення, тим компактніші групи;

− силуетний коефіцієнт (Silhouette) − міра того, наскільки об'єкт схожий на свій кластер порівняно з іншими. Значення $> 0.5$ вказує на надійну структуру.

5. Практичний кейс: Аналіз деградації земель за даними ДЗЗ

Розглянемо задачу виділення зон опустелювання. Вхідні дані − значення вегетаційного індексу NDVI та індексу вологості NDWI для 500 контрольних точок.

1. Етап попередньої обробки: вилучення точок з хмарністю та нормалізація індексів;

2. Кластеризація (DBSCAN): виявлення зон з аномально низькою вологістю та рослинністю. Використання DBSCAN дозволяє ігнорувати поодинокі висушені ділянки, фокусуючись на масивах деградації;

3. Інтерпретація: отримані кластери накладаються на векторну карту типів ґрунтів для виявлення причин деградації (ерозія, перевипас або кліматичні зміни).

Контрольні питання:

1. поясніть, чому класифікація потребує розмічених даних, а кластеризація − ні;

2. за яких умов еколог має надати перевагу алгоритму DBSCAN над K-means;

3. наведіть приклади «шумів» у геопросторових даних, які можуть спотворити результати кластеризації;

4. яким чином метод головних компонент (PCA) допомагає боротися з «прокляттям розмірності» при обробці супутникових даних;

5. опишіть фізичний зміст центроїда кластера при районуванні акваторії моря за солоністю та температурою.

Завдання для самостійної роботи: Використовуючи набір даних про рівні забруднення ґрунту (Pb, Cu, Zn), провести порівняльний аналіз розбиття території на 2 та 4 кластери. Розрахувати силуетний коефіцієнт для обох випадків та обґрунтувати оптимальне число зон екологічної напруженості.

# ЛЕКЦІЯ №5: ПОГЛИБЛЕНІ МЕТОДИ МОДЕЛЮВАННЯ, ПРОБЛЕМА ГЕНЕРАЛІЗАЦІЇ ТА АРХІТЕКТУРНА ПАРАДИГМА НЕЙРОННИХ МЕРЕЖ У ГЕОЕКОЛОГІЇ

1. Історія розвитку методів моделювання та інтелектуального аналізу

Розвиток методів, що сьогодні складають основу обробки геопросторових даних, пройшов тривалий шлях від класичної математики до сучасних нейромережевих архітектур. Розуміння історичного контексту дозволяє фахівцю з комп'ютерних наук усвідомити тяглість алгоритмічних підходів та межі їх застосування.

До ключових етапів формування дисципліни належать:

- епоха класичної апроксимації (початок XIX століття) – Карл Фрідріх Гаусс та Адрієн-Марі Лежандр незалежно один від одного розробили метод найменших квадратів (1805–1809 рр.). Спочатку метод застосовувався в астрономії для розрахунку орбіт планет, проте згодом став базовим для обробки будь-яких просторових вимірювань;
- народження терміну «регресія» (1889 рік) – Френсіс Гальтон, вивчаючи спадковість, виявив тенденцію повернення екстремальних значень до середнього показника. Це спостереження заклало підвалини для статистичного прогнозування станів складних систем;
- кібернетичний прорив та перші нейрони (1943–1958 рр.) – Воррен Маккаллок та Вальтер Піттс запропонували математичну модель біологічного нейрона. У 1958 році Френк Розенблатт створив перцептрон – першу модель нейронної мережі, здатну до навчання. Це викликало фурор, проте обмеженість тодішніх обчислювальних потужностей призвела до «першої зими штучного інтелекту»;
- ера зворотного поширення помилки (1980-ті роки) – розробка алгоритму Backpropagation Джеффрі Хінтоном та колегами дозволила

ефективно навчати багатошарові мережі, що відкрило шлях до моделювання нелінійних залежностей у природних системах;

− подолання «другої зими» ШІ (2006–2012 рр.) − впровадження методів глибокого навчання (Deep Learning) та ефективна ініціалізація ваг дозволили будувати мережі з десятками шарів;

− сучасний етап Deep Learning (з 2012 року) − перехід на обчислення за допомогою GPU та поява згорткових нейронних мереж (CNN) здійснили революцію в обробці зображень дистанційного зондування Землі (ДЗЗ), забезпечивши точність, що перевищує можливості експертного дешифрування.

Сьогодні для еколога-аналітика ці методи не є просто «чорними скриньками», а являють собою еволюційний розвиток математичного апарату апроксимації функцій у багатовимірних просторах.

2. Математична природа оптимізації в геопросторовій обробці

Будь-яка задача обробки геоданих, чи то інтерполяція рівнів забруднення, чи то класифікація підстильної поверхні, базується на парадигмі оптимізації. Метою є підбір такої моделі, яка була б максимально лаконічною та адекватною реальним спостереженням.

Математично апроксимація екологічної функції зводиться до розв'язання системи рівнянь виду $Ax = b$, де:

− $A$ − матриця вхідних ознак (наприклад, координати $X, Y$, висота над рівнем моря $Z$, значення спектральних каналів Sentinel-2);

− $x$ − вектор шуканих параметрів (ваг) моделі, які визначають вплив кожного фактора;

− $b$ − вектор реальних спостережень (концентрація токсиканта, щільність біомаси).

У геоекології ми майже завжди маємо справу з надмірними або недовизначеними системами. У таких випадках задача стає некоректно поставленою, що вимагає використання функціоналів виду:

$$argmin\left(||Ax - b||^2 + g(x)\right)$$

де $g(x)$ – член регуляризації, що обмежує складність моделі, запобігаючи її «вибуху» при обробці зашумлених польових даних. На базі цього функціоналу будуються сучасні методи Ridge та Lasso регресії, які є стандартом для попереднього аналізу екологічних факторів.

3. Проблема генералізації: Перенавчання та Недонавчання

Для спеціаліста спеціальності проблема генералізації є центральною при розробці систем моніторингу. Здатність моделі працювати на даних, яких вона не бачила під час навчання, визначає її придатність для реального використання у кризових ситуаціях.

3.1. Баланс зміщення та дисперсії (Bias-Variance Tradeoff)

- недонавчання (Underfitting): виникає при надмірному спрощенні моделі. В екології це проявляється, коли лінійна регресія ігнорує порогові ефекти токсичного впливу, вважаючи залежність суворо пропорційною;
- перенавчання (Overfitting): виникає при надмірній складності моделі відносно обсягу даних. Модель ідеально «запам'ятовує» похибки конкретних датчиків, але втрачає здатність прогнозувати стан сусідніх ділянок;
- наслідки для спеціальності: перенавчена система раннього попередження про повені може видавати хибнопозитивні сигнали через специфічні шуми гідрологічного поста, що підриває довіру до автоматизованих систем управління.

4. Специфічні архітектури нейронних мереж у геопросторовій обробці

Для вирішення сучасних екологічних задач недостатньо класичного багатошарового перцептрона (MLP). Кожен тип геопросторових даних вимагає специфічної архітектури, яка враховує внутрішню структуру інформації.

4.1. Згорткові нейронні мережі (CNN)

Призначені для обробки зображень та растрових матриць.

− архітектурна особливість: використання фільтрів (ядер згортки), що сканують зображення для виявлення локальних ознак (градієнтів, текстур, об'єктів);

− застосування в екології: архітектура U-Net для семантичної сегментації супутникових знімків. Це дозволяє попіксельно виділяти межі незаконних вирубок або контури розповсюдження амброзії у містах;

4.2. Рекурентні нейронні мережі (RNN) та архітектура LSTM

Призначені для аналізу послідовностей та часових рядів.

− архітектурна особливість: наявність зворотних зв'язків та «комірок пам'яті» (Long Short-Term Memory), що дозволяють зберігати контекст минулих станів системи;

− застосування в екології: прогнозування концентрації діоксиду азоту в атмосфері на 24 години вперед на основі ретроспективних даних метеоумов;

4.3. Графові нейронні мережі (GNN)

Обробляють дані, представлені у вигляді вузлів та зв'язків.

− архітектурна особливість: згортка виконується не по сітці пікселів, а по сусідах у графі;

− застосування в екології: аналіз мережі датчиків якості води вздовж річкового басейну, де зв'язки визначаються напрямком течії;

4.4. Трансформери (Transformers) та механізм уваги (Attention)

Сучасний стандарт для мультимодальних даних.

− архітектурна особливість: здатність моделі динамічно фокусуватися на найбільш значущих фрагментах даних незалежно від їх розташування в часі чи просторі;

− застосування в екології: архітектура Vision Transformer (ViT) для глобального кліматичного моделювання, де необхідно враховувати одночасний вплив океанічних течій у різних частинах планети;

4.5. Автоенкодери (Autoencoders)

Мережі для навчання без учителя.

− архітектурна особливість: стиснення вхідних даних у прихований простір (bottleneck) з подальшим відновленням;

− застосування в екології: виявлення аномалій (Anomaly Detection). Якщо модель навчена на «нормальному» стані довкілля, будь-який залповий викид або нетипова поведінка екосистеми призведе до високої помилки відновлення;

5. Регуляризація, Dropout та Batch Normalization: механізми стійкості

Для забезпечення стійкості прогнозів у геоекологічних задачах використовуються методи стабілізації навчання.

5.1. Метод Dropout як механізм ансамблювання

− сутність: випадкове ігнорування частини нейронів під час ітерації навчання;

− ефект: запобігає спільній адаптації нейронів, змушуючи мережу знаходити робастні ознаки. Це критично при роботі з розрідженими польовими даними, де випадкова кореляція може бути прийнята за фізичну закономірність.

5.2. Пакетна нормалізація (Batch Normalization)

− сутність: нормалізація виходів кожного шару всередині нейронної мережі;

- ефект: значно прискорює навчання та дозволяє використовувати вищу швидкість навчання (Learning Rate) без ризику розбіжності моделі при обробці мультиспектральних масивів.

6. Аугментація екологічних даних та боротьба з дефіцитом вибірок

У спеціальності дефіцит розмічених даних є головною перешкодою. Аугментація дозволяє штучно збагатити вибірку:

- спектральна аугментація: випадкова зміна яскравості та контрасту у спектральних каналах, що симулює різні умови атмосферної прозорості;
- хмарна симуляція: накладання напівпрозорих масок хмар на чисті супутникові знімки для підготовки моделі до роботи в реальних погодних умовах;
- часова аугментація: створення синтетичних станів вегетації за допомогою генеративно-змагальних мереж (GAN), що дозволяє моделювати фази росту рослин між реальними прольотами супутника.

7. Прикладні кейси застосування поглиблених моделей

- Smart City Air Monitoring: поєднання LSTM для прогнозу часових рядів та CNN для аналізу розповсюдження забруднювачів у щільній міській забудові. Це дозволяє створювати «цифрові двійники» атмосфери міста;
- Precision Agriculture (Точне землеробство): використання ViT для аналізу стану ґрунтового покриву. Модель навчається на хмарних платформах (Google Earth Engine) та передає ваги на мобільні пристрої агрономів для офлайн-роботи;
- Marine Ecology: застосування автоенкодерів для стиснення величезних масивів гідроакустичних даних, що дозволяє передавати лише значущі ознаки через супутниковий зв'язок з обмеженою пропускною здатністю.

Контрольні питання та завдання:

1. поясніть, чому згорткові мережі (CNN) є більш ефективними для аналізу супутникових знімків порівняно з MLP;
2. за яких умов архітектура LSTM є незамінною при моделюванні розливу річок;
3. опишіть роль механізму уваги (Attention) у задачах мультимодального моніторингу (текст + зображення);
4. яким чином графові нейронні мережі враховують топологію інженерних екологічних систем;
5. наведіть приклад «аугментації хмарності» та обґрунтуйте її необхідність для стабільності класифікаторів;
6. порівняйте L1 та L2 регуляризацію з точки зору їх впливу на інтерпретованість екологічної моделі;
7. поясніть фізичний зміст прихованого простору (bottleneck) в автоенкодерах при стисненні гіперспектральних кубів даних.

Завдання для самостійної роботи: Розробити концептуальну схему нейромережевої системи для моніторингу незаконних сміттєзвалищ, обґрунтувавши вибір архітектури (CNN чи ViT), методів регуляризації та стратегії аугментації. Підготувати порівняльну таблицю обчислювальної складності різних типів шарів (згорткових vs рекурентних) для впровадження на Edge-пристроях.

# ЛЕКЦІЯ №6: ІМОВІРНІСНИЙ (БАЙЄСІВСЬКИЙ) МЕТОД РОЗПІЗНАВАННЯ ТА АРХІТЕКТУРА ІМОВІРНІСНИХ НЕЙРОННИХ МЕРЕЖ У ЗАДАЧАХ ЕКОЛОГІЧНОГО МОНІТОРИНГУ

## 1. Історія розвитку імовірнісних методів у науці про дані

Імовірнісний підхід до розпізнавання образів та прийняття рішень має глибоке коріння, що сягає XVIII століття. Для фахівця з комп'ютерних наук важливо розуміти, що сучасні алгоритми класифікації геоданих базуються на фундаментах, закладених задовго до появи перших обчислювальних машин.

Ключові віхи розвитку методології:

- розробка теореми Байєса (1763 рік) – пресвітеріанський священник та математик Томас Байєс у своїй праці «Досвід розв'язання проблеми у теорії імовірностей» сформулював принцип оновлення імовірності події на основі нових даних. Цей підхід став основою для всього сучасного статистичного висновку;
- внесок П'єра-Симона Лапласа (1812 рік) – незалежно від Байєса, Лаплас розвинув імовірнісну теорію та дав їй сучасне математичне формулювання, що дозволило застосовувати ці методи у природничих науках та астрономії;
- становлення статистичної теорії розпізнавання (1950–1960 рр.) – з появою перших ЕОМ виникла потреба в алгоритмах, здатних приймати рішення в умовах невизначеності. Роботи Девіда Блеквелла та Абрахама Вальда заклали основи теорії статистичних рішень;
- поява наївного байєсівського класифікатора (1960-ті рр.) – метод почав активно використовуватися для фільтрації текстів та медичної діагностики. У геоінформаційних системах (ГІС) його почали

застосовувати для класифікації земного покриву за супутниковими знімками;

− розробка імовірнісних нейронних мереж (PNN) (1990 рік) − Дональд Спехт запропонував архітектуру нейронної мережі, що базується на байєсівській стратегії рішень та непараметричній оцінці щільності імовірності.

Сьогодні байєсівський підхід є стандартом для систем, де ціна помилки є високою (наприклад, прогнозування екологічних катастроф), оскільки він дозволяє оперувати не просто мітками класів, а ступенем впевненості у кожному рішенні.

**2. Теоретичні основи байєсівського класифікатора в геоекології**

В обробці геопросторових даних ми постійно стикаємося з імовірнісною природою явищ. Наприклад, наявність певної спектральної сигнатури на знімку не гарантує на 100%, що це хвойний ліс; вона лише робить цю гіпотезу більш імовірною порівняно з іншими.

**2.1. Формула Байєса для розпізнавання образів**

Математично задача класифікації екологічного стану $C$ за вектором ознак $X$ (наприклад, метеоумови та концентрації газів) описується формулою:

$$P(X) = \frac{P(C) \cdot P(C)}{P(X)}$$

де:

− $P(X)$ − апостеріорна імовірність (імовірність того, що об'єкт належить до класу $C$ при спостереженні ознак $X$);

− $P(C)$ − функція правдоподібності (імовірність спостерігати ознаки $X$, якщо об'єкт точно належить до класу $C$);

− $P(C)$ − апріорна імовірність (наші початкові знання про поширеність класу на даній території);

– $P(X)$ – повна імовірність спостереження вектора ознак (нормувальний множник).

**2.2. Критерій максимуму апостеріорної імовірності (MAP)**

Система обробки геоданих приймає рішення на користь того класу, для якого значення $P(X)$ є максимальним. Для еколога це означає вибір найбільш обґрунтованого діагнозу стану середовища з урахуванням історичного досвіду (апріорних даних) та поточних вимірювань.

**3. Наївний байєсівський класифікатор (NBC) у ГІС-технологіях**

Для спеціаліста NBC є одним із найшвидших та найефективніших алгоритмів для обробки великих растрових масивів.

**3.1. Гіпотеза «наївності» та її фізичний зміст**

Алгоритм називається «наївним», оскільки він базується на припущенні про незалежність ознак. Тобто імовірність того, що в точці висока температура, вважається незалежною від імовірності того, що там низька вологість (хоча фізично вони пов'язані). Незважаючи на це спрощення, NBC показує вражаючі результати в екологічному картографуванні завдяки:

– обчислювальній простоті, що дозволяє обробляти мільйони пікселів супутникових знімків за секунди;

– стійкості до нерелевантних ознак, які автоматично нівелюються в процесі множення імовірностей;

– здатності працювати з невеликими навчальними вибірками (дефіцит розмічених даних у польовій екології).

**3.2. Алгоритм роботи NBC для еколога-програміста**

1. формування навчальної матриці на основі точок наземного контролю;

2. обчислення апріорних імовірностей класів (наприклад, частка заповідних земель у загальній площі регіону);

3. оцінка параметрів розподілу (середнє та дисперсія) для кожної екологічної ознаки всередині кожного класу;

4. класифікація нових точок шляхом обчислення добутку імовірностей за всіма вимірюваними факторами.

## 4. Імовірнісні нейронні мережі (PNN) у моделюванні біосфери

PNN − це специфічний тип штучних нейронних мереж, архітектура яких безпосередньо реалізує байєсівську стратегію прийняття рішень. Вона є незамінною для задач, де необхідно миттєво адаптувати модель до нових екологічних даних.

### 4.1. Чотиришарова архітектура PNN

− вхідний шар (Input Layer) − отримує нормовані геопросторові ознаки;

− шар образів (Pattern Layer) − містить нейрони, кожен з яких представляє окремий приклад з навчальної вибірки (наприклад, спектральний профіль конкретної ділянки забруднення). Обчислюється відстань від вхідного вектора до еталона через радіальну базисну функцію;

− шар підсумовування (Summation Layer) − агрегує виходи нейронів шару образів окремо для кожного класу, формуючи оцінку щільності імовірності;

− вихідний шар (Output Layer) − порівнює отримані імовірності (з урахуванням апріорних ваг) та видає фінальну мітку екологічного стану.

### 4.2. Переваги PNN для екології

− швидкість навчання − на відміну від класичних мереж з Backpropagation, PNN «навчається» за один прохід, просто зберігаючи навчальні вектори у шарі образів;

− прозорість прийняття рішень − кожен нейрон у мережі має фізичну інтерпретацію (конкретний екологічний замір);

− стійкість до викидів − завдяки використанню ядерних функцій (Parzen window), мережа згладжує поодинокі інструментальні помилки датчиків.

**5. Чинники, що впливають на точність імовірнісної обробки**

Для досягнення 10-сторінкової глибини аналізу необхідно розглянути фактори, що визначають якість байєсівських моделей:

− якість апріорної інформації − якщо ми помилково вважаємо, що певна хвороба лісу зустрічається рідко (низьке $P(C)$), алгоритм буде ігнорувати реальні симптоми, поки вони не стануть критичними;

− кореляція ознак − у системах моніторингу повітря концентрації $CO$ та $NO_2$ часто корелюють. Наївний Байєс може «переоцінити» цей вплив, що вимагає попередньої обробки за допомогою PCA;

− обсяг навчальної вибірки − для PNN важливо мати репрезентативні приклади всіх можливих станів екосистеми, інакше «незнайомі» ситуації будуть класифіковані з низькою впевненістю;

− вибір параметра розмиття (Spread) у PNN − занадто мале значення призведе до перенавчання (мережа реагуватиме лише на точні копії еталонів), а занадто велике − до недонавчання (втрата деталізації меж між класами).

**6. Прикладні кейси застосування імовірнісних методів**

Для спеціаліста важливо бачити, як імовірнісна математика втілюється у програмний код для вирішення екологічних задач:

− **Прогнозування екологічних ризиків затоплення:** використання байєсівських мереж для оцінки імовірності виходу річки з берегів. Апріорна імовірність базується на історичних даних за 100 років, а функція правдоподібності оновлюється в реальному часі за даними датчиків рівня води та прогнозів опадів;

– **Детектування нафтових плям на морі:** імовірнісна нейронна мережа аналізує радарні знімки акваторії. Вона повинна відрізнити нафтову плівку від природних зон штилю («biogenic slicks»). PNN дозволяє задати вищу «ціну помилки» за пропуск плями, ніж за хибну тривогу;

– **Класифікація біотопів у Smart-лісництвах:** наївний байєсівський класифікатор обробляє дані з мультиспектральних камер дронів. Висока швидкість алгоритму дозволяє виконувати класифікацію безпосередньо на борту БПЛА (Edge Computing) для виявлення зон посухи;

– **Аналіз поширення інвазивних видів:** імовірнісне моделювання ареалів проживання видів (SDM – Species Distribution Modeling). Алгоритм обчислює імовірність появи інвазивної рослини у кожному квадраті географічної сітки на основі кліматичних преференцій виду.

**7. Розрахункова частина: Байєсівське районування (Приклад)**

Уявімо задачу класифікації стану ґрунту за двома ознаками ($x_1$ – вміст гумусу, $x_2$ – кислотність). Маємо два класи: $C_1$ (родючі) та $C_2$ (виснажені).

– Апріорні імовірності: $P(C_1) = 0.6, P(C_2) = 0.4$.

– Для нової проби $X$ розраховані правдоподібності: $P(C_1) = 0.15, P(C_2) = 0.30$.

– Обчислимо апостеріорні значення (без знаменника $P(X)$, оскільки він однаковий):

– $Score(C_1) = 0.15 \cdot 0.6 = 0.09$;

– $Score(C_2) = 0.30 \cdot 0.4 = 0.12$.

– Оскільки $Score(C_2) > Score(C_1)$, система автоматично маркує ділянку як виснажену, попри те, що апріорно родючих земель у регіоні більше. Це демонструє, як нові дані перекривають початкові припущення.

**Контрольні питання та завдання:**

1. поясніть роль апріорної імовірності у запобіганні помилковим екологічним прогнозам при дефіциті свіжих даних;
2. за яких умов припущення про «наївність» NBC стає критичною перешкодою для точності класифікації ландшафтів;
3. порівняйте швидкість навчання PNN та багатошарового перцептрона (MLP) при обробці динамічних потоків геоданих;
4. яким чином параметр Spread у PNN впливає на здатність мережі до генералізації екологічних знань;
5. опишіть математичний зміст функції правдоподібності у контексті детектування аномальних викидів в атмосферу;
6. розробіть схему байєсівського висновку для задачі підтвердження факту незаконного видобутку копалин за непрямими ознаками;
7. поясніть, чому імовірнісні методи є більш стійкими до «шумів» у геопросторових даних порівняно з детермінованими методами.

**Завдання для самостійної роботи:** Опрацювати методику розрахунку байєсівського класифікатора для трикласової задачі моніторингу стану лісу (здоровий, ослаблений, загиблий). Підготувати програмний код на Python з використанням GaussianNB для аналізу масиву екологічних ознак. Провести дослідження впливу апріорних ваг на фінальну точність картографування ризиків.

# ЛЕКЦІЯ №7: ЦИФРОВА ОБРОБКА АКУСТИЧНИХ СИГНАЛІВ, СПЕКТРАЛЬНИЙ АНАЛІЗ ТА БІОАКУСТИЧНИЙ МОНІТОРИНГ ДОВКІЛЛЯ

## 1. Історія розвитку систем розпізнавання звукових сигналів

Еволюція методів обробки звуку пройшла шлях від механічних пристроїв, що реагували на певні частоти, до складних нейромережевих систем, здатних ідентифікувати тисячі біологічних видів за їх вокалізацією. Для фахівця важливо розуміти, як технології розпізнавання мови стали фундаментом для сучасного екологічного моніторингу.

Ключові віхи розвитку технології:

- 1927 рік – створення робота «Герберт Телевокс» (США). Це був пристрій, де набір реле реагував на звукові сигнали певної частоти, що подавалися через телефонну лінію. Це заклало основу концепції частотної селекції сигналів;
- 1952 рік – система «Audrey» (Bell Labs). Перша спроба автоматичного розпізнавання цифр, вимовлених однією людиною. Машина займала величезну шафу і базувалася на аналізі спектральних формант;
- 1970-ті роки – поява системи «Harpy» (Університет Карнегі-Меллона), яка вперше оперувала словником понад 1000 слів. У цей період почали застосовуватися приховані марковські моделі (HMM), які досі використовуються в аналізі часових рядів екологічних даних;
- 1990-ті роки – комерціалізація систем (IBM MedSpeak), перехід від розпізнавання окремих команд до обробки безперервного потоку сигналів;
- 2010-ті роки – «глибоке навчання» (Deep Learning) та перетворення Фур’є стали стандартом для біоакустики. Поява проектів

типу *BirdNET* дозволила автоматизувати глобальний моніторинг біорізноманіття.

Сьогодні методи, що розроблялися для розпізнавання людської мови, є базою для детектування звуків лісових пожеж, незаконних вирубок (звук бензопили) або моніторингу морських ссавців в океані.

**2. Акустична екологія та цифрова репрезентація сигналів**

Звуковий сигнал у природі — це безперервна (аналогова) зміна тиску середовища. Для комп’ютерної обробки ми повинні перетворити його у дискретну форму.

**2.1. Дискретизація та квантування**

Для фахівця з комп'ютерних наук обробка починається з двох параметрів:

- частота дискретизації ($f_s$) – кількість замірів амплітуди за секунду. Згідно з теоремою Котельникова-Найквіста, для коректного відновлення сигналу частота дискретизації має бути вдвічі вищою за максимальну частоту в сигналі. В екології для ультразвукового моніторингу кажанів використовуються частоти до 192–384 кГц;
- розрядність квантування – визначає точність вимірювання амплітуди (динамічний діапазон).

**2.2. Представлення у часовій області (Time Domain)**

Це пряма залежність амплітуди від часу. Вона дозволяє визначити:

- енергію сигналу та його тривалість;
- наявність імпульсних шумів (наприклад, постріли браконьєрів або тріск дерев під час пожежі);
- періодичність повторення сигналів (цикли активності фауни).

**3. Спектральний аналіз: Перетворення Фур’є**

Більшість інформативних ознак екологічного сигналу приховані у його частотному складі. Для їх вилучення використовується перетворення Фур'є.

### 3.1. Математичне обґрунтування

Дискретне перетворення Фур'є (DFT) базується на твердженні, що будь-який складний сигнал (звук птаха, вітру чи бензопили) можна розкласти на суму простих гармонійних коливань (синусоїд та косинусоїд) різної частоти та амплітуди. Математично це перетворення сигналу з часової області у частотну виконується за формулою:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j\frac{2\pi}{N}kn}$$

де:

- $x[n]$ − вхідний відлік сигналу у часі, отриманий після дискретизації;
- $X[k]$ − комплексне число, що описує амплітуду та фазу частотної складової $k$;
- $e^{-j\frac{2\pi}{N}kn}$ − комплексний експоненціальний множник (згідно з формулою Ейлера: $\cos cos \phi - j \sin sin$ ), який виконує роль «математичного сита», що виділяє конкретну частоту з загального сигналу;
- $N$ − розмір вікна аналізу (кількість відліків, що обробляються за один раз).

Фізичний зміст цього рівняння полягає в обчисленні ступеня кореляції між вхідним сигналом та набором опорних гармонік. Якщо в сигналі присутня частота, що відповідає індексу $k$, значення $X[k]$ буде мати велику амплітуду. Таким чином, ми отримуємо «відбиток» звуку, де замість хаотичних коливань амплітуди ми бачимо чіткі спектральні піки, характерні для конкретного біологічного чи техногенного об'єкта.

### 3.2.Актуальність методу

Для спеціаліста перехід у частотну область є критичним з наступних міркувань:

- ідентифікація прихованих паттернів – у часовій області звуки різних видів птахів можуть виглядати однаково (як шум), але на спектрограмі вони мають унікальні частотні «портрети»;
- фільтрація шумів – знаючи спектральний склад фонового шуму (вітер, дощ), ми можемо математично видалити ці частоти, залишивши лише корисний сигнал (голос тварини);
- стиснення даних – у частотній області сигнал часто є розрідженим (багато частот мають нульову амплітуду), що дозволяє ефективно зберігати терабайти екологічних записів;
- автоматичне детектування аномалій – алгоритми можуть реагувати на появу специфічних частот (наприклад, низькочастотний гул техніки у заповідній зоні), що неможливо зробити аналізом лише амплітуди.

### 3.3. Швидке перетворення Фур'є (FFT) та віконні функції

Оскільки екологічні сигнали не є стаціонарними (звук птаха змінюється в часі), ми використовуємо короткочасне перетворення Фур'є (STFT). Сигнал розбивається на кадри, до кожного з яких застосовується віконна функція (наприклад, вікно Хеммінга), щоб уникнути спектральних витоків на межах кадрів.

## 4. Вилучення інформативних ознак екологічних сигналів

Для систем класифікації (наприклад, розпізнавання виду птаха) сирий спектр є занадто надмірним. Ми вилучаємо компактний вектор ознак:

- спектральний центроїд – характеризує «яскравість» звуку, вказуючи, на якій частоті зосереджена основна енергія;

- спектральний спад (Roll-off) – частота, нижче якої зосереджено 85-95% енергії спектра;
- мел-кепстральні коефіцієнти (MFCC) – враховують логарифмічне сприйняття частот, що ідеально підходить для моделювання біологічних звукових систем;
- форманти – резонансні частоти, що визначають унікальний «голос» особини або специфічний гул техногенного обладнання.

**5. Радіально-базисні нейронні мережі (RBF) у класифікації сигналів**

RBF-мережі є ефективною альтернативою багатошаровим перцептронам для задач швидкого розпізнавання акустичних сигналів.

**5.1. Архітектура мережі**

- вхідний шар – приймає вектор ознак (наприклад, набір MFCC);
- прихований (радіальний) шар – обчислює відстань між вхідним вектором та центрами «еталонних» звуків за допомогою функції Гаусса:
- $\phi(x) = e^{-\frac{|x-c|^2}{2\sigma^2}}$
- вихідний шар – лінійна комбінація сигналів з прихованого шару, що дає фінальний клас звуку.

**5.2. Переваги в екології**

- локальна апроксимація – мережа реагує лише на звуки, близькі до еталонних, що зменшує кількість хибних спрацювань на невідомі шуми;
- швидке навчання – ваги вихідного шару можна обчислити методом найменших квадратів за одну ітерацію, що дозволяє розгортати системи на автономних Edge-пристроях у заповідниках.

**6. Прикладні кейси застосування у спеціальності**

Для фахівця з комп'ютерних наук обробка сигналів − це не лише математика, а й архітектура систем:

- **Моніторинг біорізноманіття (Passive Acoustic Monitoring - PAM):** мережа автономних реєстраторів у лісі записує звук 24/7. Алгоритм FFT виділяє спектрограми, а згорткова нейромережа класифікує види комах, птахів та кажанів. Це дозволяє оцінити стан екосистеми без фізичної появи людей;
- **Боротьба з незаконною діяльністю:** детектування звуку бензопили або пострілів у реальному часі. Використання RBF-мереж дозволяє миттєво відрізнити грім або звук падіння дерева від роботи інструменту та передати координати через супутниковий зв'язок;
- **Аналіз шумового забруднення міст:** побудова просторових карт акустичного комфорту. Аналізуються частотні діапазони шуму транспорту, будівництва та промисловості для оптимізації міського планування;
- **Гідроакустичний моніторинг морів:** розпізнавання сигналів китоподібних та оцінка рівня антропогенного шуму (судноплавство, видобуток копалин), що впливає на міграції морської фауни.

**7. Практичне завдання: Побудова спектрального профілю**

В межах лабораторної роботи студенти повинні реалізувати наступний алгоритм на Python:

1. завантаження аудіосигналу (бібліотека librosa або scipy.io.wavfile);
2. преакцентна фільтрація для посилення високих частот;
3. віконне перетворення Фур’є та побудова сонограми (спектрограми);
4. класифікація сигналу за допомогою радіально-базисної моделі.

**Контрольні питання та завдання:**

1. поясніть фізичний зміст переходу від часової області до частотної при аналізі звуку бензопили;

2. яким чином теорема Котельникова визначає вимоги до апаратної частини екологічного сенсора для запису ультразвуку;
3. опишіть переваги вікна Хеммінга порівняно з прямокутним вікном при обробці коротких пташиних сигналів;
4. порівняйте ефективність RBF-мереж та логістичної регресії для задачі детектування аномальних шумів у заповідниках;
5. наведіть приклади «інформативних ознак» для розпізнавання звуку вітрогенераторів та їх впливу на локальну фауну;
6. як впливає частота дискретизації на обчислювальну складність алгоритму FFT в автономних системах моніторингу;
7. поясніть роль «формант» у задачах ідентифікації окремих особин тварин за їх голосом.

**Завдання для самостійної роботи:** Опрацювати методику перетворення аудіосигналу на набір MFCC-коефіцієнтів. Підготувати порівняльний аналіз спектрограм трьох різних видів екологічних шумів (дощ, вітер, грім). Розробити концепцію Edge-пристрою для розпізнавання сигналів тривоги тварин у зонах підвищеного антропогенного ризику.

# ЛЕКЦІЯ №8: ТЕОРІЯ ЧАСОВИХ РЯДІВ, СТАТИСТИЧНЕ ОЦІНЮВАННЯ ТА ПРЕДИКТИВНЕ МОДЕЛЮВАННЯ ДИНАМІЧНИХ ЕКОСИСТЕМ

## 1. Генеза та еволюція методів аналізу часових рядів

Аналіз часових рядів (ЧР) пройшов шлях від простих візуальних спостережень за небесними тілами до надскладних нейромережевих моделей реального часу. Для розробника систем моніторингу розуміння історії методів дозволяє обирати адекватний інструментарій відповідно до обчислювальної складності задачі.

Основні етапи розвитку:

- давньосхідна астрономія (III тис. до н.е.) – перші зафіксовані спроби виявлення періодичності у природних явищах (цикли розливу річок, сонячні цикли). Це заклало фундамент поняття циклічності;
- епоха демографічної та економічної статистики (XVII–XVIII ст.) – Вільям Петті та Джон Граунт почали використовувати впорядковані в часі дані для аналізу смертності та народжуваності, що призвело до появи методів вирівнювання рядів;
- теоретичний прорив (1920-ті роки) – Джордж Уолкер, Юл та Слуцький розробили концепції авторегресії (AR) та ковзного середнього (MA). Вони математично довели, що складні коливання можуть бути результатом випадкових імпульсів, пропущених через лінійні фільтри;
- ера Бокса-Дженкінса (1970 рік) – вихід фундаментальної праці «Аналіз часових рядів: прогноз та управління» систематизував методи ARIMA, які досі є промисловим стандартом у предиктивній аналітиці;
- ера нелінійної динаміки та хаосу (1980–1990-ті рр.) – впровадження методів розрахунку експонент Ляпунова та фрактальної розмірності для аналізу турбулентних процесів в атмосфері;

– сучасний нейромережевий етап (з 2010-х рр.) – домінування архітектур LSTM (Long Short-Term Memory) та трансформерів (Informer, Temporal Fusion Transformer), здатних обробляти мільйони паралельних екологічних часових рядів.

**2. Математичне визначення та класифікація часових рядів**

Часовий ряд – це послідовність спостережуваних значень $X = \{x_1, x_2, ..., x_t, ..., x_n\}$, впорядкованих у невипадкові моменти часу $T = \{t_1, t_2, ..., t_n\}$.

В екологічному моніторингу ЧР класифікуються за наступними ознаками:

– розмірність: одновимірні (наприклад, тільки концентрація $CO_2$) та багатовимірні або мультисерійні (сукупність показників: температура, вологість, швидкість вітру, рівень озону);

– тип часу: дискретні (заміри раз на годину) та безперервні (аналоговий запис звуку або вібрацій);

– статистична природа: детерміновані (значення точно визначаються законом) та стохастичні (містять випадкову компоненту).

Для фахівця головною метою обробки ЧР є ідентифікація природи ряду та побудова математичної моделі для екстраполяції – прогнозування майбутніх значень.

**3. Декомпозиція часового ряду: Структурні компоненти**

Будь-який складний екологічний процес можна розкласти на суму (адитивна модель) або добуток (мультиплікативна модель) чотирьох компонент:

– тренд ($T_t$) – довгострокова тенденція зміни середнього значення (наприклад, глобальне потепління за останні 100 років);

– сезонність ($S_t$) – періодичні коливання з фіксованим кроком (добові зміни температури, сезонні міграції, вегетаційні цикли);

– циклічність ($C_t$) – довгострокові коливання з нефіксованим періодом (цикли сонячної активності, багаторічні цикли посух);

– шум або випадкова компонента ($E_t$) – непередбачувані флуктуації, спричинені випадковими факторами або похибками вимірювальних приладів.

Математично адитивна модель має вигляд: $X_t = T_t + S_t + C_t + E_t$. Видалення тренду та сезонності (десезонізація) є обов'язковим кроком перед глибоким аналізом, щоб «оголити» приховані закономірності.

## 4. Стаціонарність та кореляційний аналіз у геопросторовій обробці

Для коректного застосування більшості статистичних методів ряд має бути стаціонарним (у вузькому або широкому сенсі).

### 4.1. Поняття стаціонарності

Стаціонарним називається ряд, математичне сподівання та дисперсія якого не залежать від часу, а коваріація залежить тільки від часового лагу (відстані між відліками). В екології стаціонарність – це велика рідкість, тому використовуються операції диференціювання (взяття різниць сусідніх значень).

### 4.2. Тіснота зв'язку: Коефіцієнт парної кореляції

Для аналізу взаємовпливаючих рядів (наприклад, як температура повітря впливає на споживання газу) використовується коефіцієнт кореляції Пірсона:

$$r = \frac{\Sigma\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\Sigma\left(x_i - \bar{x}\right)^2 \cdot \Sigma\left(y_i - \bar{y}\right)^2}}$$

де:

– $x_i, y_i$ – значення двох різних показників;

– $\overline{x}, \overline{y}$ – їх середні значення за період спостереження.

Інтерпретація результату для екологічного прогнозу:

– $r = 0$ – величини незалежні;

– $0 < r < 0.5$ – слабкий зв'язок;

– $0.5 \leq r < 0.8$ – зв'язок середньої сили;

– $r \geq 0.8$ – сильний кореляційний зв'язок.

**5. Методи моделювання та прогнозування**

Для розробки програмного забезпечення моніторингу використовуються різні рівні складності алгоритмів.

**5.1. Авторегресійні моделі (ARIMA)**

Модель інтегрованого ковзного середнього (AutoRegressive Integrated Moving Average) базується на припущенні, що поточне значення ряду залежить від його попередніх значень та попередніх помилок прогнозу.

– застосування: прогнозування місячних опадів або рівнів води у річках;

– перевага: математична строгість та можливість оцінки довірчих інтервалів.

**5.2. Експоненціальне згладжування**

Метод надає більшу вагу новішим даним та меншу – застарілим. Це дозволяє системі швидко адаптуватися до змін у навколишньому середовищі.

$$S_t = \alpha \cdot X_t + (1 - \alpha) \cdot S_{t-1}$$

де $\alpha$ – коефіцієнт згладжування.

**5.3. Нейромережі типу LSTM**

Для спеціаліста LSTM є пріоритетним вибором при роботі з Big Data. Завдяки механізму «ворота забування» (Forget Gate), мережа здатна

пам’ятати критичні екологічні події, що сталися сотні кроків назад, ігноруючи неважливий шум.

**6. Прикладні кейси для фахівців в екології**

- **Аналіз часових рядів біржових торгів квотами на викиди:** прогнозування вартості вуглецевих кредитів для промислових підприємств на основі глобальної економічної динаміки;
- **Моніторинг споживання енергоресурсів (Smart Grid):** аналіз ЧР енергоспоживання міста для оптимізації роботи електростанцій та зменшення вуглецевого сліду;
- **Прогнозування цвітіння водойм:** аналіз багатовимірних рядів (вміст нітратів + температура води + сонячна радіація) для випуску попереджень для рибних господарств;
- **Діагностика стану обладнання очисних споруд:** виявлення аномалій у вібраційних часових рядах насосів для запобігання аварійним витокам стічних вод.

**7. Розрахункова частина: Аналіз кореляції (Приклад)**

Уявімо задачу оцінки зв'язку між середньорічним доходом населення та обсягом товарообігу (як непрямим індикатором антропогенного тиску на ресурси). Дані за 10 років (див. таблицю 8.1):

1. розраховуємо середні значення для обох рядів;
2. обчислюємо суму добутків відхилень;
3. отримуємо значення $r = 0.96$. Висновок: між показниками існує прямий та дуже сильний зв'язок, що дозволяє будувати високоточні регресійні моделі прогнозування споживання ресурсів.

**Контрольні питання та завдання:**

1. поясніть різницю між адитивною та мультиплікативною моделями часового ряду на прикладі сезонних коливань біомаси планктону;

2. яким чином операція взяття різниць допомагає привести нестаціонарний ряд екологічних даних до стаціонарного вигляду;

3. обґрунтуйте вибір коефіцієнта згладжування α для системи моніторингу шлейфу забруднення при різких змінах напрямку вітру;

4. порівняйте ефективність ARIMA та LSTM для довгострокового прогнозування кліматичних змін;

5. опишіть фізичний зміст «автокореляції» у часових рядах спостережень за рівнем радіаційного фону;

6. як впливає частота дискретизації екологічного ряду на виявлення високочастотних компонент шуму;

7. розробіть алгоритм детектування викидів-аномалій у часових рядах сенсорної мережі Smart City.

**Завдання для самостійної роботи:** Опрацювати методику розрахунку автокореляційної функції (ACF) для ряду добових температур. Провести порівняльний аналіз точності прогнозу за допомогою простого ковзного середнього та експоненціального згладжування. Підготувати програмний сценарій на Python для декомпозиції часового ряду за допомогою бібліотеки statsmodels.

# ЛЕКЦІЯ №9: ЕВОЛЮЦІЙНІ ОБЧИСЛЕННЯ ТА ГЕНЕТИЧНІ АЛГОРИТМИ В ЗАДАЧАХ ГЛОБАЛЬНОЇ ОПТИМІЗАЦІЇ

## 1. Вступ: Природний інтелект як інженерна парадигма

Протягом мільярдів років еволюція вирішувала найскладніші задачі оптимізації, адаптуючи живі організми до постійно мінливих умов середовища. В інженерній практиці та екологічному моделюванні ми часто стикаємося з задачами, де традиційні градієнтні методи є неефективними через складність цільової функції, наявність багатьох локальних екстремумів або дискретність простору пошуку.

Еволюційні алгоритми (ЕА) – це клас евристичних методів пошуку та оптимізації, заснованих на механізмах природного відбору. Для фахівця, що працює в галузі екологічної безпеки, ці алгоритми є незамінними при вирішенні задач:

- мінімізації антропогенного навантаження при проектуванні промислових об'єктів;
- оптимізації розміщення сенсорних мереж моніторингу на складних територіях;
- управління ресурсами та логістики відходів;
- моделювання динаміки популяцій у відповідь на токсичний вплив.

Основна ідея полягає у переході від пошуку рішення однією «точкою» до роботи з «популяцією» потенційних рішень, що забезпечує глобальність пошуку та стійкість до зашумлених даних.

## 2. Історія створення та наукова генеза

Ідея використання принципів біологічної еволюції для обчислень визрівала паралельно в різних наукових школах світу.

До ключових етапів формування еволюційної парадигми належать:

− теоретичний фундамент (1859 рік) − вихід праці Чарльза Дарвіна «Походження видів», де було сформульовано принципи мінливості, спадковості та природного відбору. Це стало концептуальним ядром для майбутніх алгоритмів;

− перші цифрові експерименти (1950-ті роки) − Нільс Барічеллі та Алекс Фрейзер почали використовувати комп'ютерні симуляції для вивчення генетичних процесів. Вони довели, що навіть прості правила рекомбінації кодів можуть призводити до появи складних структур;

− еволюційне програмування (1960 рік) − Лоуренс Фогель запропонував метод оптимізації скінченних автоматів, імітуючи процес мутацій та відбору;

− еволюційні стратегії (1964 рік) − Інго Рехенберг та Ганс-Пауль Швефель розробили методи для оптимізації аеродинамічних профілів, де параметри змінювалися випадково, а виживали лише кращі варіанти;

− генетичні алгоритми (1975 рік) − Джон Холланд опублікував працю «Адаптація в природних і штучних системах», де ввів поняття хромосоми, кросинговеру та формалізував ГА як метод математичної оптимізації. Саме Холланд довів «теорему схем», яка пояснює ефективність ГА.

Сучасний етап розвитку характеризується гібридизацією ЕА з нейронними мережами та появою нових метаевристик, натхненних поведінкою комах, птахів та фізичних систем.

**3. Фундаментальна термінологія та аналогії**

Генетичний алгоритм (ГА) використовує біологічну лексику для опису математичних структур:

− особина (індивід) − потенційне рішення задачі. У задачах екологічної безпеки це може бути конкретний план розміщення очисних споруд;

− хромосома − спосіб представлення рішення у вигляді рядка (зазвичай бінарного або дійсного);

− ген − елементарна частина хромосоми, що кодує конкретний параметр (наприклад, координату датчика або потужність фільтра);

− популяція − множина особин, що існують одночасно на певній ітерації алгоритму;

− функція пристосованості (Fitness Function) − цільова функція, що визначає «якість» рішення. В екології це може бути мінімізація ризику перевищення ГДК.

**4. Механізм роботи генетичного алгоритму**

Класичний цикл ГА складається з послідовності кроків, спрямованих на поступове покращення середньої пристосованості популяції.

**4.1. Створення початкової популяції**

Алгоритм генерує набір випадкових рішень. Для забезпечення глобальності пошуку важливо, щоб початкова популяція була достатньо різноманітною та охоплювала весь допустимий простір параметрів.

**4.2. Оцінка пристосованості**

Для кожної особини розраховується значення цільової функції. Чим вище пристосованість, тим більшу імовірність має особина на передачу своїх генів наступному поколінню.

**4.3. Селекція (Відбір)**

Це процес вибору батьківських пар. Поширеними методами є:

− рулетка − імовірність вибору пропорційна пристосованості;

− турнірна селекція − вибір кращого з випадкової групи;

− ранжування − імовірність залежить від позиції особини у впорядкованому списку.

**4.4. Кросинговер (Схрещування)**

Оператор рекомбінації, що дозволяє поєднати позитивні риси двох батьківських рішень для створення нащадків. Це ключовий механізм обміну інформацією в ГА.

**4.5. Мутація**

Випадкова зміна деяких генів у нащадків. Мутація виконує роль «генератора новизни», запобігаючи передчасному сходженню алгоритму в локальний оптимум та дозволяючи досліджувати раніше невідомі області простору пошуку.

**5. Застосування в задачах екологічного моделювання та безпеки**

Еволюційні методи демонструють високу ефективність у задачах, де аналітичне рішення є неможливим:

- оптимізація мереж моніторингу: пошук мінімальної кількості станцій контролю якості повітря, які забезпечують максимальну точність картографування забруднень;
- управління відходами: розрахунок оптимальних маршрутів збору та переробки сміття з метою мінімізації паливних викидів та логістичних витрат;
- ідентифікація джерел забруднення: пошук найбільш імовірного місця та часу викиду в атмосферу на основі даних зворотного моделювання та показань сенсорів;
- проектування стійких екосистем: вибір оптимального складу рослинності для рекультивації земель, враховуючи стійкість до токсикантів та кліматичну адаптивність.

**6. Інші біоінспіровані алгоритми**

Крім ГА, в обробці геоданих активно використовуються методи ройового інтелекту:

- алгоритм мурашиної колонії (Ant Colony) – для пошуку оптимальних шляхів у графах та логістики;

− алгоритм рою частинок (Particle Swarm) − для безперервної оптимізації параметрів фізичних моделей;

− алгоритм штучної імунної системи − для виявлення аномальних викидів та кіберзагроз в екологічних мережах;

− алгоритм бджолиної колонії − для виділення найбільш перспективних ділянок при обробці супутникових знімків.

**Контрольні питання та завдання:**

1. поясніть роль мутації в запобіганні передчасній зупинці алгоритму в точці локального оптимуму;
2. яким чином вибір функції пристосованості впливає на стратегію екологічної безпеки при проектуванні промислових зон;
3. порівняйте ефективність генетичних алгоритмів та класичного градієнтного спуску для задач з розривною цільовою функцією;
4. опишіть фізичний зміст «хромосоми» у задачі оптимізації параметрів моделі поширення радіонуклідів у ґрунті;
5. наведіть приклади «екологічних обмежень», які можуть бути інтегровані в генетичний алгоритм у вигляді штрафних функцій;
6. поясніть, чому «сходження популяції» не завжди означає знаходження глобально кращого рішення;
7. розробіть концептуальну схему ГА для задачі вибору місця будівництва сміттєпереробного заводу з урахуванням рози вітрів та щільності населення.

**Завдання для самостійної роботи:** Опрацювати методику кодування параметрів за допомогою сірого коду (Gray code) для підвищення стабільності ГА. Підготувати програмний код на Python з використанням бібліотеки PyGAD для вирішення простої оптимізаційної задачі (наприклад, задача про рюкзак або пошук максимуму функції двох змінних). Провести аналіз впливу розміру популяції на час обчислень та якість отриманого екологічного прогнозу.

# ЛЕКЦІЯ №10: ТЕОРІЯ НЕЧІТКИХ МНОЖИН, НЕЧІТКА ЛОГІКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

## 1. Історія виникнення та філософські передумови нечіткої логіки

Класична математика та логіка, починаючи від Арістотеля, базувалися на принципі двозначності: твердження може бути або істинним, або хибним. Однак при описі природних систем, екологічних процесів та людського мислення ми постійно стикаємося з категоріями, які не мають чітких меж. Поняття «забруднена вода», «тепла погода» або «високий ризик» неможливо адекватно описати за допомогою лише 0 та 1.

Ключові етапи розвитку теорії:

- передумови в античності та середньовіччі – ще античні філософи зауважували парадокси «купи» (скільки піщинок утворюють купу?), що вказувало на розмитість людських понять;
- праці Яна Лукасевича (1920-ті рр.) – польський логік розробив багатозначні логіки, де твердження могли мати проміжні значення істинності;
- фундаментальна робота Лотфі Заде (1965 рік) – професор Каліфорнійського університету в Берклі опублікував статтю «Fuzzy Sets» у журналі «Information and Control». Саме Заде формалізував математичний апарат, який дозволив комп'ютерам оперувати нечіткими поняттями;
- розвиток нечітких контролерів (1970-1980-ті рр.) – Ебрахім Мамдані (Ebrahim Mamdani) застосував нечітку логіку для керування паровим двигуном. В цей же період японські корпорації почали масово впроваджувати нечіткі алгоритми в побутову техніку та системи управління транспортом (наприклад, метрополітен у місті Сендай);

− сучасний етап (з 1990-х рр.) − інтеграція нечіткої логіки з нейронними мережами (ANFIS) та її широке застосування в екологічному моделюванні, медичній діагностиці та аналізі геопросторових даних.

**2. Від даних до знань: Експертні системи в екологічній безпеці**

Для фахівця в галузі екологічної безпеки критично важливим є перехід від обробки числових даних до оперування знаннями. Якщо дані − це результати вимірювань (наприклад, концентрація азоту), то знання − це правила їх інтерпретації (наприклад, «якщо азоту багато, а кисню мало − екосистемі загрожує евтрофікація»).

Перші системи, засновані на знаннях (knowledge-based systems), з'явилися у 1970-х роках:

− MYCIN − система для діагностики інфекційних захворювань крові;

− DENDRAL − система для розшифровки даних мас-спектрометрії в хімії.

В екології нечіткі експертні системи дозволяють акумулювати досвід провідних екологів та автоматично застосовувати його для оцінки ризиків у режимі реального часу, навіть якщо вхідна інформація є неповною або сумнівною.

**3. Метод експертних оцінок у формуванні нечітких моделей**

Метод експертних оцінок − це системна процедура отримання інформації від висококваліфікованих фахівців (експертів) для аналізу об’єктів, характеристики яких не можуть бути виміряні прямими інструментальними методами. У геоекологічних системах цей метод є незамінним для формалізації якісних ознак середовища.

Основними завданнями експертного оцінювання в межах нечіткої логіки є:

− визначення структури лінгвістичних змінних;

− побудова функцій приналежності для кожного терму (наприклад, визначення меж «високої температури»);

− формування бази правил («ЯКЩО... ТО...»);

− визначення вагових коефіцієнтів факторів екологічного ризику.

До найбільш поширених методів збору експертної інформації належать:

− метод анкетування та інтерв'ювання;

− метод «Дельфі» − багаторазове анонімне опитування з ознайомленням експертів із результатами попереднього туру для досягнення консенсусу;

− метод парних порівнянь − попарне зіставлення факторів для обчислення їхньої відносної значущості;

− метод мозкового штурму − колективне генерування ідей у неструктурованому форматі.

Отримана експертна інформація піддається статистичній обробці для перевірки узгодженості думок (наприклад, за допомогою коефіцієнта конкордації Кендалла). Якщо думки експертів узгоджені, дані використовуються для налаштування нечіткої системи.

**4. Математичні основи нечітких множин**

Нечітка множина $A$ на універсальній множині $X$ визначається як сукупність пар $\left(x, \mu_A(x)\right)$, де $\mu_A(x)$ − функція приналежності, яка ставить у відповідність кожному елементу число з інтервалу $[0, 1]$.

**4.1. Функції приналежності (Membership Functions)**

Функція приналежності відображає суб'єктивну або статистичну впевненість у тому, що елемент належить до певної категорії. В обробці геоданих найчастіше використовують:

− трикутні та трапецієподібні функції − зручні для швидких обчислень у реальному часі;

− гаусові функції − краще описують природні розподіли екологічних факторів;

− сигмоїдальні функції − використовуються для опису переходів між станами (наприклад, «безпечно» − «небезпечно»).

**4.2. Операції над нечіткими множинами**

Для нечіткої логіки визначені аналоги класичних операцій:

− перетин (логічне «І») − зазвичай обчислюється як мінімум значень функцій приналежності;

− об'єднання (логічне «АБО») − обчислюється як максимум значень;

− доповнення (логічне «НЕ») − обчислюється як $1\mu_A(x)$.

5. Лінгвістичні змінні та правила нечіткого висновку

Лінгвістична змінна − це змінна, значеннями якої є слова або словосполучення природної мови. Наприклад, змінна «Рівень забруднення» може приймати значення: «Низький», «Середній», «Високий», «Критичний».

Система нечіткого висновку базується на правилах виду «ЯКЩО <умова>, ТО <висновок>». В екологічній безпеці це виглядає так:

− «ЯКЩО концентрація пилу Висока І швидкість вітру Низька, ТО рівень екологічного ризику Дуже Високий».

− Процес отримання результату складається з чотирьох етапів:

− фазифікація − перетворення чітких значень датчиків у ступені істинності нечітких термів;

− нечіткий висновок − застосування бази правил для визначення результуючої нечіткої множини;

− композиція − об'єднання результатів усіх правил, що спрацювали;

– дефазифікація – перетворення отриманої нечіткої множини у конкретне числове значення (наприклад, бал ризику від 0 до 100). Найпоширенішим методом є метод центру мас (centroid).

**6. М'яка кластеризація: Алгоритм Fuzzy C-Means (FCM)**

Традиційна кластеризація (k-means) жорстко відносить точку до одного кластера. Однак екосистеми не мають різких кордонів. Перехідна зона між лісом та лугом (екотон) містить характеристики обох типів.

М'яка кластеризація дозволяє кожній точці належати до всіх кластерів з певним ступенем впевненості $w_{ij} \in [0, 1]$. Сума ступенів належності для кожної точки завжди дорівнює одиниці.

Алгоритм FCM ітераційно мінімізує функціонал:

$$J_m = \sum_{i=1}^{n} \sum_{j=1}^{c} w_{ij}^{m} |x_i - \beta_j|^2$$

де $m > 1$ – показник розмитості (fuzziness). Цей метод є незамінним при районуванні територій зі складним градієнтом забруднення, де неможливо провести чітку межу між зонами впливу різних джерел.

**7. Прикладні кейси в екологічній безпеці та моніторингу**

Застосування нечіткої логіки дозволяє вирішувати задачі, які важко піддаються строгому математичному опису:

– комплексне оцінювання якості поверхневих вод – об'єднання десятків хімічних показників у єдиний лінгвістичний індекс стану водойми;

– прогноз пожежної небезпеки – врахування нечітких факторів: «сухість підстилки», «віддаленість від доріг», «імовірність перебування людей»;

− управління процесами очищення стічних вод − нечіткі контролери дозволяють підтримувати оптимальний рівень реагентів в умовах постійної зміни вхідного потоку;

− класифікація земного покриву за супутниковими знімками − розпізнавання об'єктів з розмитими межами (болота, рідколісся, зони підтоплення).

**Контрольні питання та завдання:**

1. поясніть основну відмінність між імовірністю та нечіткістю при описі стану довкілля;

2. за яких умов доцільно використовувати нечітку кластеризацію замість класичного методу k-середніх;

3. опишіть етапи функціонування системи нечіткого висновку на прикладі датчика диму;

4. яким чином вибір форми функції приналежності (трикутна vs гаусова) впливає на чутливість екологічної моделі;

5. наведіть приклади лінгвістичних змінних, які є критичними для забезпечення екологічної безпеки морських портів;

6. поясніть фізичний зміст параметра «розмитості» $m$ в алгоритмі Fuzzy C-Means;

7. розробіть спрощену базу з трьох нечітких правил для оцінки ризику ерозії ґрунтів залежно від нахилу поверхні та щільності рослинності.

Завдання для самостійної роботи: Опрацювати методику дефазифікації за методом центру мас для заданої нечіткої множини результатів. Підготувати реферат на тему «Порівняльний аналіз систем нечіткого висновку Мамдані та Сугено в задачах кліматичного прогнозування». Провести програмну симуляцію нечіткого контролера для стабілізації рівня розчиненого кисню у ставку (використовуючи Python бібліотеку scikit-fuzzy).

# ЛЕКЦІЯ №11: АЛГОРИТМИ НЕЧІТКОГО ВИСНОВКУ МАМДАНІ ТА СУГЕНО: МАТЕМАТИЧНИЙ АПАРАТ ТА ПРИКЛАДНІ АСПЕКТИ ЕКОЛОГІЧНОГО МОДЕЛЮВАННЯ

## 1. Концепція нечіткого висновку як інструменту прийняття рішень

Системи нечіткого висновку (СНВ) є вищим щаблем обробки геопросторових даних, оскільки вони дозволяють не просто накопичувати інформацію, а автоматизувати процес логічного міркування. В умовах екологічної безпеки, де вхідні дані часто є суперечливими, неточними або отриманими з різних джерел (космічні знімки, наземні пости, експертні звіти), СНВ стає «містком» між числовими вимірюваннями та управлінськими діями.

Для фахівця в галузі екологічної безпеки СНВ вирішує наступні критичні завдання:

- агрегація різнорідних показників забруднення у єдиний індекс небезпеки;
- моделювання складних нелінійних залежностей без необхідності знання точних диференціальних рівнянь процесу;
- врахування людського фактору та інтуїції експертів-екологів у автоматизованих системах контролю;
- стійкість до збоїв окремих датчиків за рахунок лінгвістичної обробки сигналів.

## 2. Історія розвитку алгоритмів нечіткого висновку

Розвиток СНВ пройшов шлях від теоретичних розробок Лотфі Заде до масового впровадження у промислові та екологічні системи управління.

Ключові віхи:

− 1973 рік − Лотфі Заде опублікував працю «Основи нового підходу до аналізу складних систем та процесів прийняття рішень», де ввів поняття лінгвістичної змінної та нечітких правил продукцій;

− 1974–1975 рр. − Ебрахім Мамдані (Ebrahim Mamdani) розробив перший нечіткий контролер для управління паровою машиною. Це довело практичну придатність нечіткої логіки для динамічних систем;

− 1985 рік − Мічіо Сугено та Такагі запропонували модель нечіткого висновку, де висновок правила є не нечіткою множиною, а лінійною функцією від вхідних змінних. Це значно спростило обчислення та відкрило шлях до адаптивних нейро-нечітких мереж (ANFIS);

− 1990-ті роки − широке впровадження нечіткої логіки в екологічне картографування та системи раннього попередження про цунамі та паводки в Японії та країнах ЄС.

Сьогодні ці моделі є стандартом де-факто для інтелектуальних датчиків «Smart City» та систем підтримки прийняття рішень у надзвичайних ситуаціях.

### 3. Алгоритм Мамдані: Класична схема нечіткого висновку

Алгоритм Мамдані є найбільш інтуїтивно зрозумілим, оскільки і умови, і висновки в ньому представлені нечіткими лінгвістичними термами.

#### 3.1. Етапи функціонування СНВ Мамдані

Процес розрахунку екологічного показника (наприклад, ризику викиду) включає:

− фазифікація − перетворення чітких значень $x$ (наприклад, концентрація газу 5.2 $мг/м^3$) у ступені істинності $\mu(x)$ для відповідних термів («Середній», «Високий»);

− агрегація (активація) — обчислення істинності умов кожного правила. Зазвичай використовується операція $min$ (для «І») та $max$ (для «АБО»);

− акумуляція − об’єднання результатів усіх правил, що спрацювали, у єдину нечітку множину для вихідної змінної;

− дефазифікація − перетворення підсумкової «фігури» (множини) у чітке числове значення.

### 3.2. Математичне обґрунтування методів дефазифікації

Для отримання чіткої рекомендації (наприклад, на скільки відсотків потрібно зменшити викиди) використовують методи:

− метод центру мас (Centroid) − найбільш робастний метод, що враховує всю форму нечіткої множини:

$$z = \frac{\int \mu(z) z dz}{\int \mu(z) dz}$$

− метод медіани (Bisector) − точка, що ділить площу фігури навпіл;

− методи максимумів (SOM, LOM, MOM) − вибір найменшого, найбільшого або середнього значення серед тих, що мають максимальну істинність. Використовуються, коли критично важливо знати пікове значення ризику.

## 4. Алгоритм Сугено: Обчислювальна ефективність та адаптивність

У моделі Сугено висновок правила є поліномом від вхідних змінних. Типове правило першого порядку має вигляд: « ЯКЩО $x \in A$ І $y \in B$, ТО $z = p \cdot x + q \cdot y + r$».

Переваги моделі Сугено в екологічній безпеці:

− відсутність етапу дефазифікації у класичному розумінні − замість інтегрування виконується просте середньозважене обчислення;

− висока швидкість обробки даних у реальному часі для великих територій;

− можливість автоматичного налаштування параметрів $p, q, r$ за допомогою методів машинного навчання на основі історичних даних моніторингу.

**5. Порівняльна характеристика моделей для задач моніторингу**

Для вибору адекватної моделі фахівець має враховувати наступні чинники:

− модель Мамдані: краще підходить для експертних систем, де важливо зберегти семантику природної мови та зрозумілість правил для еколога;

− модель Сугено: краще підходить для систем автоматичного регулювання (наприклад, очисні споруди) та задач апроксимації поверхонь забруднення;

− стійкість до завад: обидві моделі демонструють значну гнучкість при роботі з зашумленими вхідними даними, проте Мамдані забезпечує більш плавний вихідний сигнал.

**6. Прикладні кейси застосування в екологічній безпеці**

Розглянемо сценарії реалізації СНВ для забезпечення екологічної безпеки територій:

− **Оцінка якості питної води:** вхідні змінні − «Вміст нітратів», «Мутність», «pH». База знань містить 27 правил, сформованих експертами-гідробіологами. Результат − індекс придатності води, що візуалізується на ГІС-картах у реальному часі;

− **Система раннього попередження про лісові пожежі:** СНВ аналізує нечіткі дані про «Температуру», «Швидкість вітру» та «Вологість підстилки». Навіть при невеликих перевищеннях порогів система може

видати статус «Висока імовірність», що дозволяє мобілізувати ресурси до появи відкритого вогню;

- **Управління очищенням стічних вод:** нечіткий контролер регулює подачу повітря в аеротенки залежно від концентрації кисню та біологічного споживання кисню (БСК). Це дозволяє економити до 30% електроенергії, забезпечуючи при цьому нормативну якість очистки;
- **Картографування шумового навантаження аеропортів:** модель Сугено використовується для інтерполяції рівнів шуму між точками замірів, враховуючи «Тип літака» та «Час доби» як параметри полінома.

**7. Практичне завдання: Розрахунок СНВ для оцінки ризику**

В межах лабораторної роботи необхідно розв’язати систему нечіткого висновку для задачі оцінки стану ґрунту:

1. задати вхідні змінні $A$ (рівень гумусу) та $D$ (кислотність);
2. сформувати базу з 3-5 правил продукцій;
3. виконати фазифікацію для конкретного заміру;
4. провести акумуляцію та дефазифікацію методом центру мас;
5. проаналізувати чутливість моделі до зміни форми функцій приналежності.

**Контрольні питання та завдання:**

1. поясніть математичну роль операції $min$ при агрегації умов у правилах Мамдані;
2. за яких умов у задачах екологічної безпеки доцільно використовувати метод дефазифікації LOM (Largest of Maximum);
3. порівняйте архітектури Мамдані та Сугено з точки зору складності їхньої реалізації на мові Python;
4. яким чином використання нечіткого висновку допомагає уникнути «стрибків» у прийнятті рішень при граничних значеннях концентрацій забруднювачів;

5. опишіть фізичний зміст коефіцієнтів лінійної функції у висновках правил моделі Сугено;

6. як впливає кількість термів у лінгвістичній змінній на точність екологічного прогнозу та обчислювальну складність системи;

7. розробіть нечітке правило для системи безпеки морського терміналу, що враховує імовірність розливу нафтопродуктів.

**Завдання для самостійної роботи:** Опрацювати алгоритм побудови СНВ типу Мамдані для трикритеріальної оцінки екологічного стану річки. Підготувати порівняльну таблицю методів дефазифікації за критеріями плавності та обчислювальної інтенсивності. Виконати програмну реалізацію моделі Сугено для апроксимації функції залежності рівня шуму від дистанції до автостради.

# ЛЕКЦІЯ №12: РЕКУРЕНТНІ НЕЙРОННІ МЕРЕЖІ: МЕРЕЖА ХОПФІЛДА ТА ДВОСПРЯМОВАНА АСОЦІАТИВНА ПАМ'ЯТЬ У ЗАДАЧАХ ЕКОЛОГІЧНОГО МОНІТОРИНГУ

## 1. Історія виникнення та наукова генеза рекурентних мереж

Розвиток нейромережевих архітектур, здатних до запам'ятовування та відновлення образів, став одним із найважливіших етапів у становленні штучного інтелекту. На відміну від мереж прямого поширення, рекурентні мережі мають зворотні зв'язки, що дозволяє їм зберігати «внутрішній стан» або пам'ять.

Ключові етапи розвитку:

- теорія Дональда Гебба (1949 рік) – канадський нейропсихолог сформулював правило, згідно з яким зв’язки між нейронами зміцнюються, якщо вони активуються одночасно. Це правило («нейрони, що розряжаються разом, зв’язуються разом») стало математичною основою для навчання асоціативної пам’яті;
- робота Джона Хопфілда (1982 рік) – фізик Джон Хопфілд запропонував архітектуру нейронної мережі з повнозв’язною топологією та довів, що вона має властивості енергетичної мінімізації. Це дозволило розглядати процес розпізнавання образів як рух системи до стану стійкої рівноваги (атрактора);
- двоспрямована асоціативна пам'ять Барта Коско (1988 рік) – Коско розширив ідеї Хопфілда, запропонувавши модель BAM (Bidirectional Associative Memory), яка дозволяє асоціювати образи з різних просторів ознак (гетероасоціація);
- розвиток теорії атракторів та хаосу (1990-ті рр.) – використання рекурентних мереж для моделювання складних екологічних циклів та нелінійної динаміки атмосферних процесів.

Сьогодні ці архітектури є фундаментом для створення систем, здатних відновлювати зашумлені супутникові знімки або ідентифікувати складні екологічні патерни навіть при частковій втраті даних з датчиків.

**2. Концепція асоціативної пам'яті та її типи**

Асоціативна пам'ять − це здатність системи відновлювати повний зразок інформації на основі його фрагмента або зашумленої копії. У фахівця з екологічної безпеки це асоціюється з процесом «добудови» цілісної картини екологічної ситуації при обмежених ресурсах спостереження.

Розрізняють три основні класи асоціативної пам'яті:

− автоасоціативна пам'ять − система відновлює еталонний образ за його спотвореною копією (вхідний та вихідний простори співпадають). Наприклад, відновлення повної карти забруднення акваторії за фрагментарними даними з декількох буїв;

− гетероасоціативна пам'ять − система встановлює зв'язок між двома різними образами. Наприклад, асоціація певного спектрального профілю супутникового знімка з конкретним типом токсичного забруднення ґрунту;

− інтерполяційна пам'ять − здатність системи видавати середньозважений результат при отриманні сигналу, що знаходиться між двома вивченими еталонами.

**3. Дискретна мережа Хопфілда: Архітектура та навчання**

Мережа Хопфілда − це одношарова рекурентна нейронна мережа, де кожен нейрон з'єднаний з усіма іншими нейронами, крім самого себе.

**3.1. Математична модель нейрона**

Стан кожного нейрона $x_i$ є біполярним ( $\{-1, 1\}$). Вихід нейрона визначається через порогову функцію активації:

$$x_i(t+1) = sign\left(\sum_{j=1}^{n} w_{ij} x_j(t) - \theta_i\right)$$

де:

- $w_{ij}$ – вага зв'язку між $j$\$ – м та \$$i$-м нейронами;
- $\theta_i$ – поріг активації (зазвичай рівний нулю).

**3.2. Алгоритм навчання (за правилом Гебба)**

Для запам'ятовування набору з $M$ еталонних образів $X^m$, матриця ваг $W$ обчислюється як сума зовнішніх добутків векторів:

$$W = \sum_{m=1}^{M} \left(X^m\right) \cdot \left(X^m\right)^T$$

При цьому діагональні елементи матриці $w_{ii}$ встановлюються рівними нулю, щоб уникнути самозбудження нейронів.

**3.3. Енергетична функція та стабільність**

Для мережі Хопфілда визначається функція Ляпунова (функція енергії):

$$E = -\frac{1}{2}\sum_i \sum_j w_{ij} x_i x_j$$

Процес функціонування мережі — це ітераційне оновлення станів нейронів, яке завжди призводить до зменшення загальної енергії системи. Стан локального мінімуму енергії відповідає одному з еталонних образів, збережених у пам'яті.

**4. Двоспрямована асоціативна пам'ять (BAM)**

Модель Барта Коско дозволяє встановлювати асоціації між векторами різної розмірності. Це двошарова мережа зі зворотними зв'язками між шарами.

**4.1. Механіка передачі сигналу**

Сигнал ітераційно передається від шару $A$ до шару $B$ і назад:

1. $B(t + 1) = sign(W \cdot A(t))$;
2. $A(t + 1) = sign\left(W^T \cdot B(t + 1)\right)$. Цей процес триває до повної стабілізації обох векторів. В екології це може використовуватися для кореляції між супутниковими даними (шар A) та результатами наземних хімічних аналізів (шар B).

**5. Застосування в задачах екологічної безпеки**

Рекурентні архітектури мають унікальні переваги для систем екологічного моніторингу:

− відновлення пошкоджених геопросторових даних: усунення «білих плям» на картах, спричинених хмарністю або збоями в передачі телеметрії;

− розпізнавання «екологічних архетипів»: автоматична ідентифікація типових сценаріїв розвитку техногенних аварій за першими непрямими ознаками;

− фільтрація шумів у сенсорних мережах: мережа Хопфілда може виступати як інтелектуальний фільтр, що відсікає аномальні похибки датчиків, повертаючи їх до найбільш імовірного стаціонарного стану;

− асоціативні бази знань: швидкий пошук аналогічних історичних ситуацій для прийняття оперативних рішень у кризових штабах.

**6. Розрахунковий приклад: Формування асоціативної матриці**

Припустимо, необхідно запам'ятати екологічний зразок (наприклад, профіль забруднення трьох зон) $X = [-1, 1, -1]$.

1. Обчислюємо матрицю ваг:

   $W = X \cdot X\hat{}T = [\![\{bmatrix\} -1 \backslash\backslash 1 \backslash\backslash -1]\!]\{bmatrix\} \cdot [\![\{bmatrix\} -1$

2. Встановлюємо діагональ в нуль:

$W = [\![\{bmatrix\} 0 \& -1 \& 1 \backslash\backslash -1 \& 0 \& -1 \backslash\backslash 1 \& -1 \& 0]\!]\{bmatrix\}$

3. Перевірка: при подачі зашумленого сигналу $[-1, -1, -1]$, мережа за одну ітерацію обчислить $W \cdot X_{noisy} = [0, 2, 0]$, що після порогової функції $sign$ поверне оригінальний образ $[-1, 1, -1]$.

**7. Обмеження та особливості проектування**

Незважаючи на потужність, мережі Хопфілда мають ряд обмежень:

− ємність пам'яті: мережа може стабільно зберігати лише близько $0.14 \cdot N$ образів (де $N$ − кількість нейронів), інакше виникає проблема перехресних завад;

− помилкові стани (спінові стекла): поява в енергетичному ландшафті додаткових мінімумів, які не відповідають жодному еталону;

− вимога до ортогональності: образи мають бути максимально несхожими (ортогональними) для забезпечення чіткого розділення центрів притягання атракторів.

**Контрольні питання та завдання:**

1. поясніть фізичний зміст енергетичної функції Хопфілда в контексті стабілізації екологічної моделі;
2. за яких умов у мережі виникають «помилкові атрактори» та як вони впливають на достовірність екологічного прогнозу;
3. порівняйте процеси автоасоціації та гетероасоціації на прикладі системи моніторингу лісових масивів;
4. обґрунтуйте використання біполярного кодування $\{-1, 1\}$ замість бінарного ($\{0, 1\}$) для покращення роздільної здатності мережі;
5. опишіть роль правила Гебба у формуванні довгострокової пам'яті інтелектуальних систем безпеки;
6. яким чином ітераційний процес у мережі BAM забезпечує збіжність до стабільного рішення при неповних вхідних даних;
7. розробіть схему використання мережі Хопфілда для верифікації цілісності даних у мережі автономних метеостанцій.

**Завдання для самостійної роботи:** Виконати розрахунок матриці ваг для мережі Хопфілда, що запам'ятовує два ортогональні образи розмірністю 4. Провести симуляцію відновлення одного з образів при інверсії одного біта. Підготувати порівняльний аналіз обчислювальної ефективності мережі Хопфілда та мережі Хеммінга для задачі класифікації гідрологічних об'єктів.

# ЛЕКЦІЇ №13–14: ТЕОРЕТИЧНІ ОСНОВИ КОМП'ЮТЕРНОГО ЗОРУ: ВІД ОБРОБКИ ЗОБРАЖЕНЬ ДО АВТОМАТИЗОВАНОГО РОЗПІЗНАВАННЯ ЕКОЛОГІЧНИХ ОБ'ЄКТІВ

## 1. Вступ та історія розвитку комп'ютерного зору

Комп'ютерний зір (Computer Vision, CV) – це міждисциплінарна галузь штучного інтелекту, що займається створенням систем, здатних отримувати, обробляти, аналізувати та «розуміти» цифрові зображення або відеодані для прийняття обґрунтованих рішень. Для фахівця з обробки геопросторових даних комп'ютерний зір є «очима» автоматизованих систем моніторингу, що дозволяють замінити суб'єктивний людський аналіз точними математичними алгоритмами.

### 1.1. Історична генеза дисципліни

Розвиток комп'ютерного зору пройшов шлях від простих експериментів із геометричними фігурами до надскладних нейромережевих архітектур:

- етап зародження (1960-ті роки) – Ларрі Робертс у 1963 році продемонстрував можливість вилучення 3D-інформації з 2D-зображень геометричних блоків, що заклало основу аналізу країв та контурів;
- концепція Девіда Марра (1970-ті роки) – запропоновано ієрархічну модель зору, де обробка йде від первинного «ескізу» (лінії, межі) до повної об'ємної моделі сцени;
- поява перших детекторів ознак (1980-ті роки) – розробка алгоритмів детектування кутів (Harris Corner Detector) та країв (Canny Edge Detector). У 1980 році Куніхіко Фукусіма запропонував «неокогнітрон» – прообраз сучасних згорткових нейромереж;

− ера інваріантних ознак (1990-ті – 2000-ні роки) − розробка алгоритму SIFT (David Lowe, 1999) дозволила розпізнавати об’єкти незалежно від масштабу, освітлення та кута зйомки;

− сучасний етап (з 2012 року) − перемога архітектури AlexNet у конкурсі ImageNet ознаменувала початок домінування Deep Learning. Сьогодні CV базується на глибоких згорткових мережах та трансформерах (Vision Transformers).

**2. Цифрова репрезентація та математика зображень**

Для комп’ютера зображення − це не візуальна сцена, а багатовимірна матриця (тензор) числових значень, організована у вигляді регулярної сітки дискретних елементів.

**2.1. Моделі представлення даних**

Фундаментальною одиницею цифрового зображення є піксель (picture element), який характеризується координатами у просторі матриці $(x, y)$ та значенням інтенсивності (яскравості) або кольору. У комп’ютерному зорі використовують різні моделі представлення цих даних, що розрізняються за глибиною дискретизації та способом кодування колірної інформації.

**Розрядність та глибина кольору**

Глибина кольору визначає динамічний діапазон зображення − кількість можливих станів (відтінків), які може приймати кожен піксель:

− 8-бітні зображення (Grayscale) − дозволяють кодувати 256 рівнів сірого (0 − абсолютно чорний, 255 − абсолютно білий). Це стандарт для більшості базових алгоритмів обробки контурів;

− 24-бітні зображення (TrueColor) − використовують по 8 біт на кожен із трьох основних каналів (RGB), що дає понад $16$ мільйонів колірних комбінацій;

− 16- та 32-бітні зображення (HDR/Floating point) − застосовуються у супутниковому моніторингу та науковій візуалізації для збереження надвисокої деталізації даних про сонячну радіацію або теплові потоки, де точність вимірювання фізичних величин є критичною.

**Порівняльний аналіз колірних просторів**

Вибір моделі представлення кольору критично впливає на ефективність сегментації екологічних об'єктів:

− модель RGB (Red, Green, Blue) − аддитивна модель, орієнтована на апаратну реалізацію (монітори, камери). Головний недолік полягає у сильній взаємозалежності колірних каналів та яскравості. При зміні освітленості (наприклад, через хмарність на супутниковому знімку) значення у всіх трьох каналах змінюються нерівномірно, що робить RGB непридатним для надійної сегментації об'єктів за кольором;

− модель HSV (Hue, Saturation, Value) − перцептивно орієнтована модель, що розділяє інформацію про колірний тон (Hue), насиченість (Saturation) та яскравість (Value). У задачах екологічного моніторингу HSV є кращою, оскільки дозволяє виділити «зелений колір» рослинності в каналі Hue, при цьому результат сегментації буде стійким до тіней та перепадів яскравості, які зосереджені в каналі Value;

− модель Lab − апаратно-незалежний простір, що описує колір через яскравість (L) та два колірнорізницеві компоненти (a, b). Модель розроблена для максимально точної відповідності людському зору та є ефективною при пошуку тонких спектральних відмінностей у стані лісових масивів.

**2.2. Математичні операції над зображеннями**

Фундаментом цифрової обробки зображень є операція згортки (Convolution), яка є процесом обчислення нового значення цільового пікселя шляхом зваженого підсумовування значень яскравості його сусідів. Математично це виражається формулою:

$$g(x,y) = \sum_i \sum_j f(x - i, y - j) \cdot h(i,j)$$

Дана операція дозволяє інтерпретувати зображення через призму системного аналізу, де компоненти формули мають наступний фізичний зміст:

– $f(x,y)$ – вхідний сигнал або функція яскравості вихідного зображення в координатах $(x,y)$;

– $h(i,j)$ – імпульсна характеристика системи або ядро фільтра (матриця ваг), що визначає характер перетворення (розмиття, підкреслення меж, пошук текстур);

– $g(x,y)$ – вихідний сигнал або результуюче значення пікселя після застосування оператора фільтрації.

Суть операції полягає у послідовному переміщенні «ковзного вікна» ядра $h$ по всьому масиву даних $f$. У кожній позиції обчислюється сума добутків відповідних елементів матриці зображення та ядра. Цей метод є основою для вилучення локальних ознак, необхідних для детекції екологічних аномалій, таких як нафтові плями або вогнища деградації рослинності.

**3. Детальний опис базових операцій обробки**

Для ефективного екологічного моніторингу необхідно розуміти фізичний та математичний зміст кожної операції, що застосовується до кадрів відеопотоку або супутникових знімків.

**3.1. Точкові операції**

Точкові операції змінюють значення кожного пікселя незалежно від його сусідів. Це найшвидші методи обробки, спрямовані на корекцію гістограми зображення та виділення значущих діапазонів даних.

**Порогова бінаризація (Thresholding)**

Це процес відображення безперервного або багаторівневого діапазону яскравостей у бінарну множину $\{0,\ 255\}$. Пікселі, чия яскравість вища за поріг $T$, стають білими, решта − чорними. В екології виділяють:

− глобальну бінаризацію − застосування єдиного порогу до всього знімка. Ефективна при високому контрасті об’єкта та фону;

− адаптивну бінаризацію − обчислення локального порогу для кожної області зображення. Незамінна при обробці супутникових знімків із нерівномірним освітленням або наявністю тіней від хмар;

− метод Оцу (Otsu) − автоматичний пошук порогу, що мінімізує дисперсію всередині класів фону та об’єкта, базуючись на аналізі гістограми інтенсивностей.

**Гістограмні перетворення**

Гістограма − це дискретний розподіл імовірностей рівнів яскравості пікселів. Операції над нею включають:

− вирівнювання гістограми (Histogram Equalization) − перерозподіл інтенсивностей для досягнення рівномірного розподілу. Дозволяє «розтягнути» контраст у зображеннях із низькою деталізацією, роблячи приховані деталі видимими;

− контрастне обмеження (CLAHE) − локальне вирівнювання гістограми, яке запобігає посиленню шуму в однорідних областях, що є критичним при аналізі текстури водної поверхні або хмарності.

**Гамма-корекція**

Нелінійне перетворення яскравості за законом $I_{out} = I_{in}$. При $\gamma < 1$ зображення стає світлішим, проявляючи деталі в тінях. При $\gamma > 1$ зображення затемнюється, підкреслюючи деталі у пересвічених областях, що корисно при аналізі відблисків на воді або снігового покриву.

**3.2. Локальні (фільтруючі) операції**

Локальні операції базуються на аналізі околиці пікселя. Вони дозволяють проводити згладжування шумів або, навпаки, детектувати різкі зміни сигналу (межі).

**Лінійна фільтрація (Розгладжування)**

- розмиття за Гауссом (Gaussian Blur) – використовує ядро, значення якого спадають від центру до країв за законом нормального розподілу. Це «фізично коректний» спосіб розмиття, який ефективно видаляє високочастотний шум, зберігаючи загальну структуру об’єктів та природні градієнти ландшафту;
- усереднюючий фільтр (Box Blur) – замінює значення пікселя середнім арифметичним його сусідів. Працює швидко, але може створювати артефакти у вигляді розмитих меж, що небажано при точному визначенні контурів об’єктів.

**Нелінійна фільтрація**

- медіанний фільтр – вибирає медіанне значення яскравості у вікні. Це один із найважливіших інструментів у екологічній безпеці, оскільки він ідеально справляється з «імпульсним» шумом (битими пікселями сенсорів або перешкодами при передачі), не розмиваючи при цьому різкі межі об’єктів;
- білатеральний фільтр – виконує розмиття з урахуванням не тільки відстані між пікселями, а й різниці їх яскравостей. Це дозволяє згладжувати шум всередині областей (наприклад, всередині лісового масиву), повністю зберігаючи чіткість його меж та контурів окремих дерев.

**Детектування меж (Диференціальні фільтри)**

- оператор Собеля (Sobel) – обчислює наближене значення градієнта яскравості по горизонталі та вертикалі. Дозволяє виділити контури об’єктів, напрямок яких збігається з осями координат, що корисно для пошуку доріг або лінійних споруд;

− оператор Лапласа (Laplacian) − обчислює другу похідну яскравості. Вкрай чутливий до шуму, але дозволяє виявити точки найрізкішої зміни сигналу, що використовується для пошуку дрібних об’єктів (наприклад, окремих суден чи будівель) на однорідному фоні.

### 3.3. Морфологічні операції

Морфологія працює з формою об’єктів, використовуючи структуруючий елемент (ядро):

− ерозія (Erosion) − «роз’їдання» меж об’єкта. Якщо у вікні ядра є хоча б один фоновий піксель, центральний піксель стає фоновим. Дозволяє прибрати дрібний шум та роз’єднати об’єкти, що доторкаються;

− дилатація (Dilation) − «розширення» меж. Якщо у вікні є хоча б один піксель об’єкта, центральний стає частиною об’єкта. Допомагає заповнити дірки та об’єднати фрагментовані частини одного образу;

− відкриття (Opening) − послідовна ерозія, а потім дилатація. Видаляє дрібні об’єкти, зберігаючи форму великих;

− закриття (Closing) − послідовна дилатація, а потім ерозія. Усуває дрібні дірки всередині об’єктів та згладжує контури.

## 4. Розширені приклади застосування в екологічній безпеці

Системи комп’ютерного зору дозволяють автоматизувати рутинні задачі, критично важливі для збереження природного балансу та запобігання катастрофам.

### 4.1. Автоматизоване сортування відходів

Застосування CV на сміттєпереробних заводах дозволяє ідентифікувати типи матеріалів (пластик, скло, папір, метал) на конвеєрній стрічці:

− спектральний аналіз − використання гіперспектральних камер для визначення хімічного складу полімерів;

− детекція контурів − визначення форми пляшок, коробок та інших ємностей для їх захоплення робоманіпулятором;

− класифікація − нейронні мережі розділяють потік відходів, мінімізуючи людський фактор та підвищуючи частку переробки.

**4.2. Моніторинг нафтових забруднень акваторій**

Ідентифікація розливів нафти на морській поверхні потребує поєднання декількох методів CV:

− текстурний аналіз − нафтова плівка змінює шорсткість поверхні води, що детектується як зміна локальних бінарних шаблонів (LBP);

− колірна сегментація в HSV − виділення райдужних розводів, характерних для тонких шарів нафтопродуктів;

− морфологічне замикання − об'єднання розрізнених плям у єдину зону забруднення для розрахунку загального об'єму викиду.

**4.3. Оцінка стану лісових ресурсів (NDVI та сегментація)**

Комп'ютерний зір у поєднанні з ДЗЗ дозволяє проводити аудит лісів:

− розрахунок вегетаційних індексів − попіксельна арифметика каналів (червоного та ближнього інфрачервоного) для оцінки фотосинтетичної активності;

− семантична сегментація (U-Net) − розділення території на класи: «здоровий ліс», «вирубка», «гар», «дорога»;

− детекція змін (Change Detection) − порівняння знімків за різні періоди для автоматичного виявлення незаконних рубок площею навіть у кілька дерев.

**4.4. Контроль популяцій та біорізноманіття**

Системи моніторингу на базі CV у заповідниках забезпечують безконтактний облік фауни:

− фотопастки з ШІ – активація запису тільки при виявленні біологічного об’єкта, що економить заряд батареї та місце на носії;

− ідентифікація особин – розпізнавання унікальних патернів забарвлення (смуги у тигрів, плями у леопардів) для відстеження міграцій;

− термальний зір – детекція теплокровних тварин у густій рослинності або в нічний час.

**5. Алгоритми сегментації та детектування**

Для виділення об’єктів у CV використовуються спеціалізовані підходи:

− алгоритм вододілу (Watershed) — сегментація, заснована на представленні зображення як топографічної карти, де яскравість пікселя – це висота;

− каскади Хаара – використання набору прямокутних фільтрів для швидкого пошуку об’єктів (наприклад, морських суден або будівель);

− YOLO (You Only Look Once) – сучасна архітектура, що дозволяє одночасно локалізувати (знайти рамку) та класифікувати об’єкт за один прохід нейронної мережі.

**Контрольні питання та завдання:**

− опишіть механізм роботи фільтра Гаусса та його роль у зниженні шумів на супутникових знімках;

− у чому полягає різниця між ерозією та дилатацією при обробці бінарних масок забруднень;

− як використання колірного простору HSV допомагає у задачах виділення водних об’єктів;

− поясніть принцип роботи алгоритма Оцу для автоматичного вибору порогу бінаризації;

− наведіть приклад використання морфологічного «закриття» для аналізу фрагментованих лісових масивів;

− яким чином нейронні мережі типу YOLO можуть бути адаптовані для детекції браконьєрських суден;

− розробіть послідовність операцій (пайплайн) CV для виявлення дикої тварини на зашумленому відео з фотопастки.

**Завдання для самостійної роботи:** Розробити на мові Python (використовуючи OpenCV) скрипт для виявлення зелених насаджень на аерофотознімку. Скрипт має включати перехід у простір HSV, застосування колірного фільтра, морфологічну чистку шумів та відмальовування контурів знайдених областей. Підготувати порівняльну таблицю ефективності фільтрів (медіанного, Гаусса та усереднюючого) для очищення зображень від туману.

# ЛЕКЦІЯ №15: ПІДСУМКИ КУРСУ ТА СТРАТЕГІЧНІ НАПРЯМИ РОЗВИТКУ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ В ЕКОЛОГІЧНІЙ БЕЗПЕЦІ

## 1. Ретроспективний аналіз курсу: від сирих даних до екологічних знань

Завершення курсу «Методи обробки геопросторових даних» дозволяє нам поглянути на вивчений матеріал як на цілісну екосистему алгоритмів, спрямованих на вирішення глобальної мети – забезпечення екологічної безпеки та сталого розвитку. Протягом семестру ми пройшли шлях еволюції методів, який повторює історію розвитку обчислювального інтелекту: від класичної статистики до глибоких нейромережевих архітектур.

Фундаментальна парадигма, на якій базувався наш курс, – це трансформація «сирих» вимірювань у дієві стратегічні рішення через ієрархію DIKW (Data, Information, Knowledge, Wisdom). Ми з'ясували, що:

- дані самі по собі є лише статичною фіксацією стану, дискретними значеннями, що потребують первинної верифікації на наявність інструментальних похибок;
- інформація з’являється тоді, коли дані отримують просторово-часову прив'язку та контекст, перетворюючись на структуровані масиви у межах ГІС-технологій;
- знання формуються шляхом виявлення прихованих закономірностей, детермінованих та стохастичних зв'язків за допомогою алгоритмів класифікації, кластеризації та регресійного аналізу;
- мудрість (або розуміння) – це вищий щабель обробки, що реалізується через здатність моделювати сценарії «що, якщо...», оцінювати ризики та приймати превентивні рішення, що мінімізують антропогенний тиск.

Для фахівця з екологічної безпеки вивчені методи є не просто набором програмних засобів, а фундаментальною математичною мовою, якою розмовляє сучасна природа у цифрову епоху.

## 2. Систематизація та поглиблення вивчених методологічних блоків

Давайте проведемо критичний огляд основних технологічних модулів, які складають аналітичне ядро дисципліни.

### 2.1. Попередня обробка та математична верифікація даних

Ми обґрунтували, що якість будь-якого екологічного прогнозу на 80% залежить від підготовки вхідної інформації. Ключовими операціями, без яких неможливе подальше навчання моделей, є:

- масштабування та нормалізація (Min-Max, Z-score), що дозволяють привести показники з різною фізичною природою до єдиного безрозмірного простору;
- виняток середнього та центрування ознак, що є критичним для коректної роботи алгоритмів лінійної алгебри та нейронних мереж;
- спектральна фільтрація та усунення шумів, зокрема використання швидкого перетворення Фур'є (FFT) для декомпозиції сигналів та виділення корисних гармонік у біоакустиці.

### 2.2. Статистичне та ймовірнісне моделювання в умовах невизначеності

Байєсівський підхід став для нас інструментом роботи з неповними даними. Ми зрозуміли, що імовірнісна природа екосистем не допускає жорстких детермінованих висновків. Апостеріорна імовірність дозволяє оцінити ступінь впевненості у наявності забруднення, враховуючи як нові заміри, так і історичний досвід (апріорні знання).

### 2.3. Машинне навчання: Класифікація, Кластеризація та Гібридизація

Ми опанували методи «цифрового районування» територій:

− кластеризація (K-means, DBSCAN, FCM) дозволила нам виявляти природні кордони екосистем та зон впливу без втручання оператора;

− класифікація (SVM, дерева рішень, випадкові ліси) дала можливість автоматично маркувати стан об'єктів за складними векторами ознак;

− ансамблеві методи об’єднали результати декількох моделей для підвищення робастності екологічного моніторингу в умовах зашумленого середовища.

### 2.4. Предиктивне моделювання та аналіз динамічних процесів

Регресійний аналіз та аналіз часових рядів стали нашими інструментами «екологічного передбачення»:

− моделі ARIMA та експоненціальне згладжування дозволяють стабільно прогнозувати стаціонарні та квазістаціонарні процеси;

− рекурентні нейронні мережі (RNN) та архітектури LSTM забезпечують можливість моделювання «пам'яті» екосистеми, що є критичним для прогнозування повеней та тривалого накопичення токсикантів у ґрунтах.

### 2.5. Інтелектуальний синтез: Нечітка логіка та Еволюційні обчислення

Ми розглянули підходи, що імітують біологічні процеси:

− нечітка логіка (алгоритми Мамдані та Сугено) формалізувала якісні описи експертів, перетворивши «розмиті» знання на чіткі інженерні коефіцієнти;

− генетичні алгоритми вирішили задачі глобальної оптимізації, де традиційні методи безсилі: пошук оптимальних маршрутів логістики відходів та конфігурації сенсорних мереж.

### 2.6. Комп’ютерний зір як метод дистанційної діагностики

Блок комп’ютерного зору завершив архітектуру курсу, додавши можливість автоматизованого візуального аналізу. Обробка зображень у просторах RGB та HSV, морфологічні операції та згорткові мережі (CNN) стали базою для ідентифікації незаконних вирубок, класифікації видів фауни та детектування пожеж на ранніх стадіях.

**3. Стратегічні тренди 2024-2026 у геопросторовій обробці**

Світ технологій розвивається у напрямку гіперзв’язності та автономності. Основними векторами, що визначатимуть екологічну безпеку, є:

- **Цифрові двійники (Digital Twins) планетарного масштабу:** інтеграція фізичних моделей з потоками даних у реальному часі для створення динамічних реплік екосистем. Це дозволяє здійснювати прогностичне управління територіями через симуляцію катастрофічних сценаріїв;
- **Explainable AI (XAI) в екології:** перехід від «чорних скриньок» до інтерпретованих моделей. Для забезпечення екологічної безпеки критично важливо розуміти, чому нейронна мережа прийняла рішення про критичний стан об'єкта;
- **Green AI та енергоефективні обчислення:** розробка алгоритмів, що потребують мінімальних енергетичних ресурсів для навчання, що саме по собі є кроком до зменшення вуглецевого сліду IT-інфраструктури;
- **Гіперспектральний моніторинг та хмарні обчислення:** масове використання платформ типу Google Earth Engine для паралельної обробки терабайтів спектральних даних, що дозволяє бачити хімічний склад атмосфери в режимі «live»;
- **Edge Intelligence:** впровадження легковажних моделей ШІ (TensorFlow Lite, TinyML) безпосередньо на автономні датчики, що

забезпечує миттєву реакцію на локальні екологічні загрози без доступу до мережі;

− **Мультимодальна аналітика:** одночасний аналіз супутникових знімків, звукових сигналів (біоакустика) та текстових звітів для створення 360-градусної картини стану довкілля.

**4. Етичні виклики та системні ризики цифровізації екології**

Впровадження інтелектуальних систем несе не лише можливості, а й нові загрози, які ми повинні враховувати:

**Алгоритмічна упередженість** використання моделей, навчених на обмежених географічних вибірках, може призвести до хибних рішень у специфічних локальних екосистемах;

**Безпека екологічних даних** захист систем моніторингу від кібератак, спрямованих на підробку показників забруднення з метою приховування злочинної діяльності;

**Техногенна залежність** надмірна довіра до автоматизованих систем може призвести до втрати навичок експертного польового аналізу та критичного мислення у фахівців;

**Екологічний слід IT-сектору** усвідомлення того, що збереження даних та навчання великих моделей вимагає значних водних та енергетичних ресурсів.

**5. Роль фахівця у новій архітектурі екологічної безпеки**

Незважаючи на домінування алгоритмів, людина залишається архітектором смислів. У новій реальності фахівець має поєднувати ролі дата-саєнтиста, еколога та системного інженера.

Завданнями нового покоління фахівців є:

− проектування гібридних інтелектуальних систем, що поєднують фізичні закони та методи машинного навчання;

− синтез знань з різних галузей для вирішення трансграничних екологічних проблем;

- забезпечення достовірності та цілісності інформаційних ланцюгів у системах моніторингу;
- розробка адаптивних стратегій управління ризиками на основі результатів предиктивного моделювання.

**6. Фінальний висновок**

Курс «Методи обробки геопросторових даних» надав вам не лише інструменти, а й методологічну базу для системного аналізу планетарних процесів. Ми довели, що математика – це найефективніший спосіб захисту природи. Світ стає складнішим, дані – масивнішими, а виклики – агресивнішими, але тепер ви володієте методами, що дозволяють перетворити ентропію інформації на структуру безпеки.

Майбутнє екологічної безпеки – це симбіоз біологічного інтелекту та обчислювальної потужності. І саме вам належить визначати межі цього симбіозу, захищаючи баланс між технологічним прогресом та цілісністю біосфери.

**Підсумкові контрольні питання до всього курсу:**

- поясніть, як поєднуються методи FFT та RBF-мережі у системах біоакустичного моніторингу заповідних зон;
- обґрунтуйте вибір між моделями Мамдані та Сугено для управління автоматизованими станціями очистки стічних вод;
- яким чином використання методів Dropout та регуляризації впливає на здатність екологічної моделі до генералізації знань;
- наведіть приклад задачі, де еволюційні алгоритми є єдиним можливим методом пошуку глобального оптимуму екологічного проекту;
- опишіть роль морфологічних операцій «відкриття» та «закриття» у пайплайнах комп’ютерного зору для аналізу супутникових знімків лісів;
- як впливає частота дискретизації часового ряду на достовірність виявлення сезонних компонент у кліматичних моделях;

− які архітектурні особливості мереж LSTM роблять їх придатними для прогнозування поширення шлейфів забруднення.

# СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ


1. Бернштейн Д. С. Еволюційні обчислення та генетичні алгоритми: теорія та практика. Київ : Наукова думка, 2018. 345 с.

2. Bishop C. M. Pattern Recognition and Machine Learning. New York : Springer, 2006. 738 p.

3. Box G. E. P., Jenkins G. M., Reinsel G. C., Ljung G. M. Time Series Analysis: Forecasting and Control. 5th ed. Hoboken : Wiley, 2015. 712 p.

4. Cressie N. Statistics for Spatial Data. Revised ed. New York : Wiley, 2015. 928 p.

5. Geographic Information Systems and Science / P. A. Longley, M. F. Goodchild, D. J. Maguire, D. W. Rhind. 4th ed. New York : Wiley, 2015. 480 p.

6. Глибоке навчання / Я. Гудфеллоу, Й. Бенджіо, А. Курвілль ; пер. з англ. Київ : ДМК прес, 2018. 652 с.

7. Гонсалес Р., Вудс Р. Цифрова обробка зображень ; пер. з англ. Москва : Техносфера, 2012. 1104 с.

8. Google Earth Engine: A Cloud-based Platform for Multi-temporal Environmental Monitoring : [електронний ресурс]. URL: https://earthengine.google.com/ (дата звернення: 28.01.2026).

9. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York : Springer, 2009. 745 p.

10. An Introduction to Statistical Learning: with Applications in R / G. James, D. Witten, T. Hastie, R. Tibshirani. New York : Springer, 2013. 426 p.

11. Макелютіс С. Геопросторова статистика та методи інтерполяції в екології. Лондон : Springer, 2021. 280 с.

12. Marr D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. San Francisco : W. H. Freeman, 1982. 397 p.

13. Mitchell T. M. Machine Learning. New York : McGraw-Hill Education, 1997. 432 p.

14. Національні та міжнародні стандарти моніторингу довкілля та вимоги до звітності з екологічної безпеки. Київ : Міністерство екології та природних ресурсів України, 2022. 156 с.

15. OpenCV: Open Source Computer Vision Library : [електронний ресурс]. URL: https://opencv.org/ (дата звернення: 28.01.2026).

16. Scikit-fuzzy: Fuzzy Logic Toolkit for SciPy : [електронний ресурс]. URL: https://pythonhosted.org/scikit-fuzzy/ (дата звернення: 28.01.2026).

17. Sentinel Hub: Processed Satellite Imagery for GIS and Land Monitoring : [електронний ресурс]. URL: https://www.sentinel-hub.com/ (дата звернення: 28.01.2026).

18. The Scientist and Engineer’s Guide to Digital Signal Processing / S. W. Smith : [електронний ресурс]. URL: https://www.dspguide.com/ (дата звернення: 28.01.2026).

19. Statsmodels: Statistical Modeling and Hypothesis Testing in Python : [електронний ресурс]. URL: https://www.statsmodels.org/ (дата звернення: 28.01.2026).

20. Zadeh L. A. Fuzzy Sets. *Information and Control*. 1965. Vol. 8, Is. 3. P. 338–353.