Міністерство освіти і науки України

Західноукраїнський національний університет

Факультет комп’ютерних інформаційних технологій

Кафедра інформаційно-обчислювальних систем і управління

Звіт

Про виконання лабораторної роботи №2

з дисципліни « Методи та системи штучного інтелекту»

Виконав:

Студент групи КНШІ-31

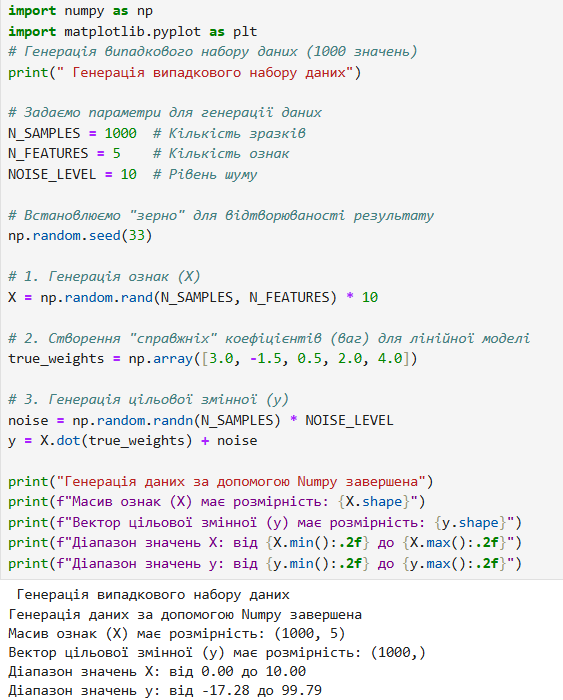
Семенюк Руслан

Тернопіль 2025

**Мета:** отримати навики з аналізу даних з використанням регресора за методом k-найближчих сусідв (knn)

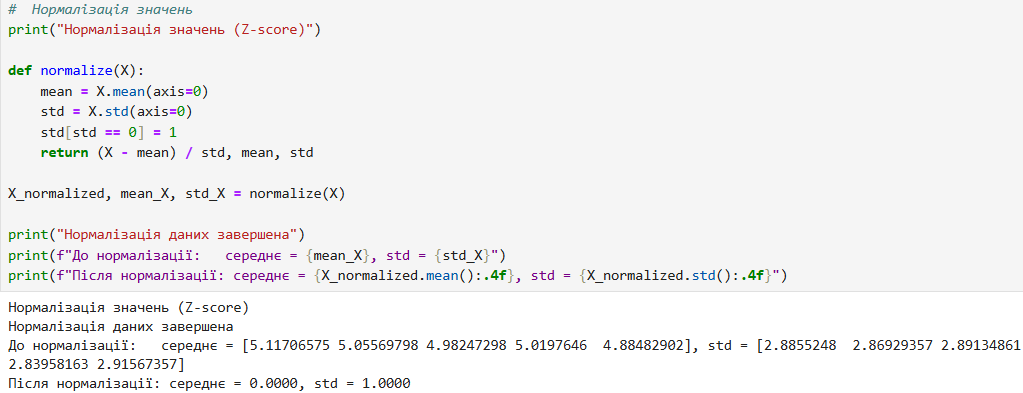
**Хід роботи**

**Крок1**. Генерація та підготовка даних



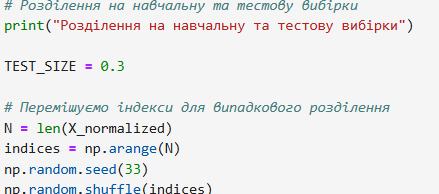
Штучно генеруємо 1000 зразків даних. Кожен зразок має 5 незалежних ознак (X) та відповідне цільове значення (y), яке ми повинні навчитися прогнозувати. Встановлено "зерно" випадковості (seed=33). Яке гарантує унікальність відтворюваність

**Крок 2**. Нормалізуємо значення



Проводиться стандартизація даних методом Z-score. Для кожної характеристики віднімається середнє значення та ділиться на стандартне відхилення. Після цієї операції всі ознаки починають мати середнє значення близьке до 0 та стандартне відхилення близьке до 1. Це важливо, оскільки алгоритм KNN використовує відстані між точками, і без нормалізації ознаки з більшими числовими значеннями мали б непропорційно великий вплив на результат.

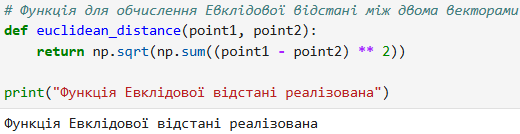
**Крок 3**: Розділення вибірок та Функція відстані





Розділяються 1000 спостережень на дві частини у співвідношенні 70%/30%. Навчальна вибірка містить 700 спостережень і використовується для "навчання" моделі - знаходження найближчих сусідів. Тестова вибірка містить 300 спостережень і слугує для незалежної оцінки якості моделі.

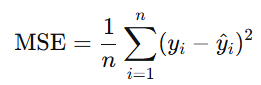
**Крок 4**: Реалізація KNN-регресора та Навчанння.



Алгоритм складається з трьох основних функцій:

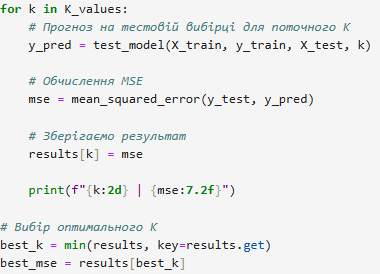
1. Функція евклідової відстані - обчислює "відстань" між точками в просторі ознак
2. Функція пошуку найближчих сусідів - знаходить k найближчих точок до заданої
3. Функція прогнозування - обчислює середнє значення цільової змінної серед знайдених сусідів

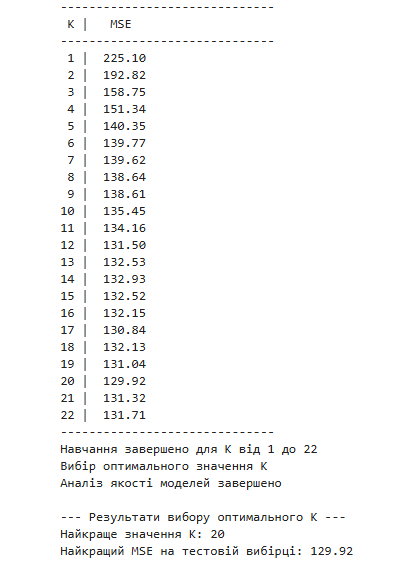
**Крок 5**: Оцінка якості для різних K та вибір найкращого

Для оцінки якості регресії ми використаємоMSE. Чим менше MSE, тим

краща модель.





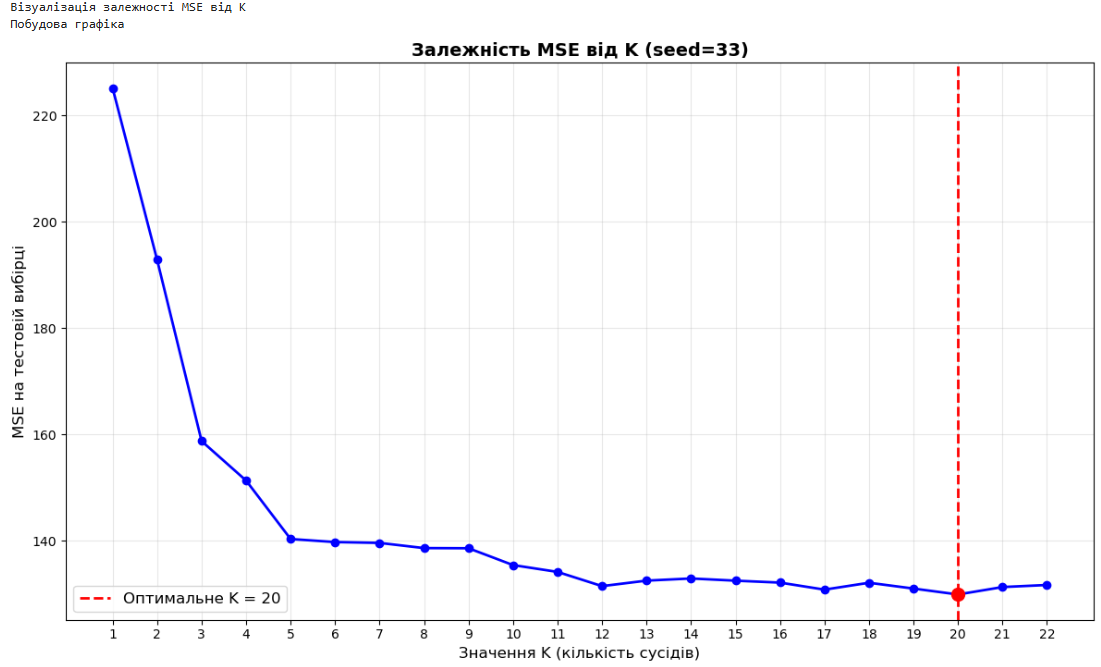


Проводиться серія експериментів, тестуючи модель з різною кількістю сусідів - від 1 до 22. Для кожного значення K:

1. Робляться прогнози для всіх 300 тестових спостережень
2. Обчислюється середня квадратична помилка (MSE) між прогнозованими та реальними значеннями
3. Фіксуються результати для подальшого аналізу

Аналізуються результати всіх експериментів для виявлення оптимального значення K, при якому модель демонструє найкращу якість прогнозування.

**Крок 6**: Візуалізація результатів



Графік показує, що оптимальна якість досягається при K=20, що свідчить про те, що для цих даних модель потребує відносно великої кількості сусідів для точних прогнозів, і underfitting починається лише за межами цього значення.

Висновок: У ході роботи було реалізовано повний цикл побудови та оптимізації KNN-регресора: згенеровано синтетичний набір даних з 1000 спостережень, проведено нормалізацію ознак, виконано розділення на навчальну та тестову вибірки у співвідношенні 70/30, реалізовано власну версію алгоритму k-найближчих сусідів з функціями евклідової відстані та прогнозування. Проведено експерименти з 22 значеннями параметра K, виявлено оптимальне значення K=20, при якому досягається мінімальна середня квадратична помилка, що підтверджено візуалізацією залежності якості моделі від кількості сусідів.

**GitHub**<https://github.com/s1pliy/AI/blob/main/Lab2/Lab2_MSAI.ipynb>