МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Західноукраїнський національний університет

Факультет комп'ютерних інформаційних технологій

Кафедра інформаційно-обчислювальних

систем і управління

Звіт про виконання лабораторної роботи №1

з дисципліни «Методи та системи штучного інтелекту»

студент групи КН-31

Ковальковський Віталій

Тернопіль – 2023

ЛІНІЙНА РЕГРЕСІЯ: АПРОКСИМАЦІЯ, ІНТЕРПОЛЯЦІЯ ТА ЕКСТРАПОЛЯЦІЯ ДАНИХ   
Завдання:

1. Завантажити дані із датасету «diabets\_world.csv»:   
% Зчитування даних з CSV-файлу

data = readmatrix('Diabets\_World.csv');

% Обмеження даних до рядків від 392 до 402

year = data(392:402, 4);

data\_values = data(392:402, 5);

2. Розділити вибірку на дві частини (дослідну і тестову) – представити графічно:   
% Розмір дослідної вибірки (наприклад, 80%)

train\_size = 0.8;

num\_data\_points = length(year);

num\_train\_points = round(train\_size \* num\_data\_points);

% Розділення на дослідну і тестову вибірку

train\_year = year(1:num\_train\_points);

train\_data\_values = data\_values(1:num\_train\_points);

test\_year = year(num\_train\_points+1:end);

test\_data\_values = data\_values(num\_train\_points+1:end);

3. Знайти точку центру мас дослідної вибірки даних (відобразити точку на графіку):

% Розрахунок точки центру мас для дослідної вибірки

center\_of\_mass\_x = mean(train\_year);

center\_of\_mass\_y = mean(train\_data\_values);

% Графічне представлення

figure;

scatter(train\_year, train\_data\_values, 60, 'b', 'filled');

hold on;

scatter(test\_year, test\_data\_values, 60, 'r', 'filled');

hold on;

% Відображення точки центру мас

scatter(center\_of\_mass\_x, center\_of\_mass\_y, 100, 'g', 'filled');

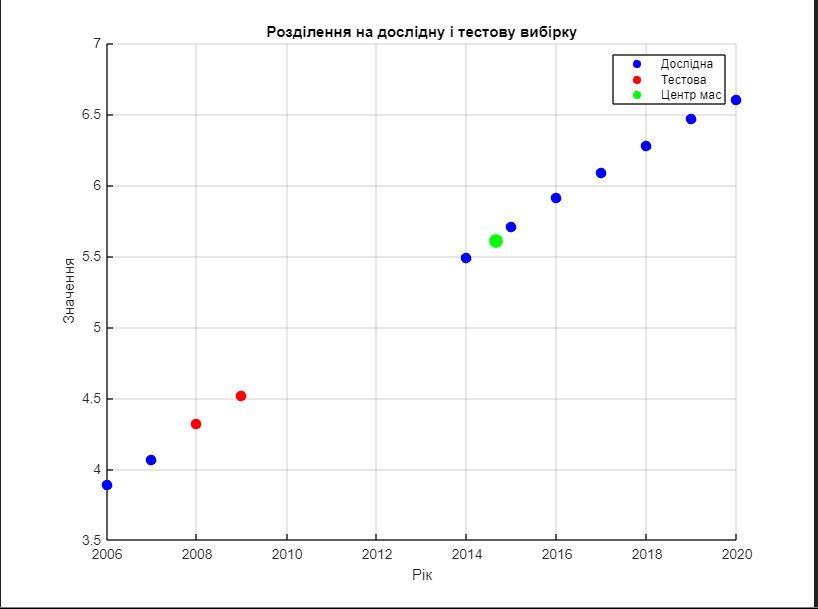
xlabel('Рік');

ylabel('Значення');

title('Розділення на дослідну і тестову вибірку');

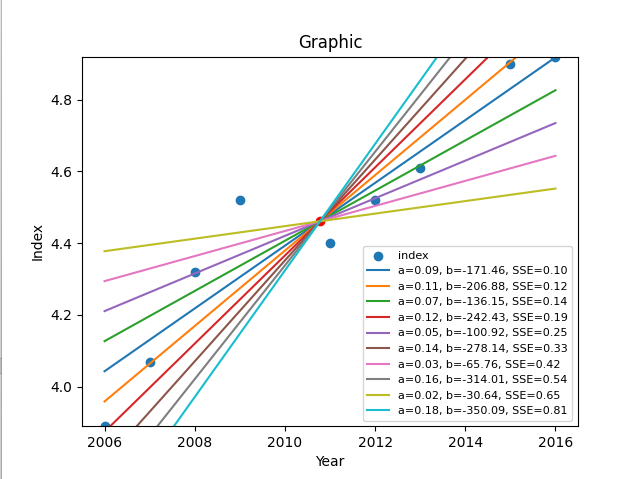
legend('Дослідна', 'Тестова', 'Центр мас');

grid on;





4. Дослідити лінійні регресії, що найкраще (з найменшими відхиленнями) відображають множину даних дослідної вибірки:



5. Забезпечити пошук та вибір лінійної регресії із найменшою сумарною похибкою між даними датасету, а також побудованими лінійними моделями (позицію відобразити у таблиці та графічно на області дослідних даних датасету):

% Вивести результати аналізу регресії та похибки

disp('Результати аналізу регресії:');

results\_table = table(slope\_values', intercept\_values', error\_values', 'VariableNames', {'Коефіцієнт\_нахилу', 'Зсув', 'Сумарна\_похибка'});

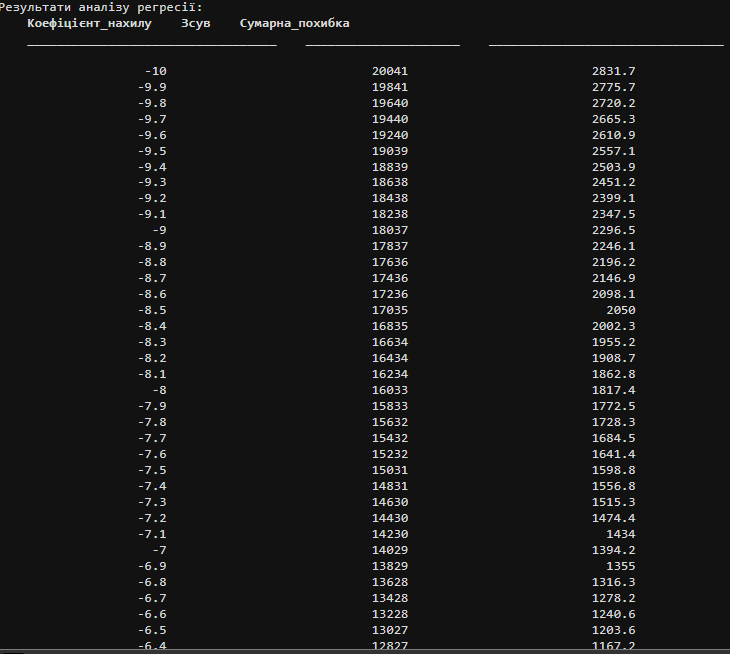
disp(results\_table);

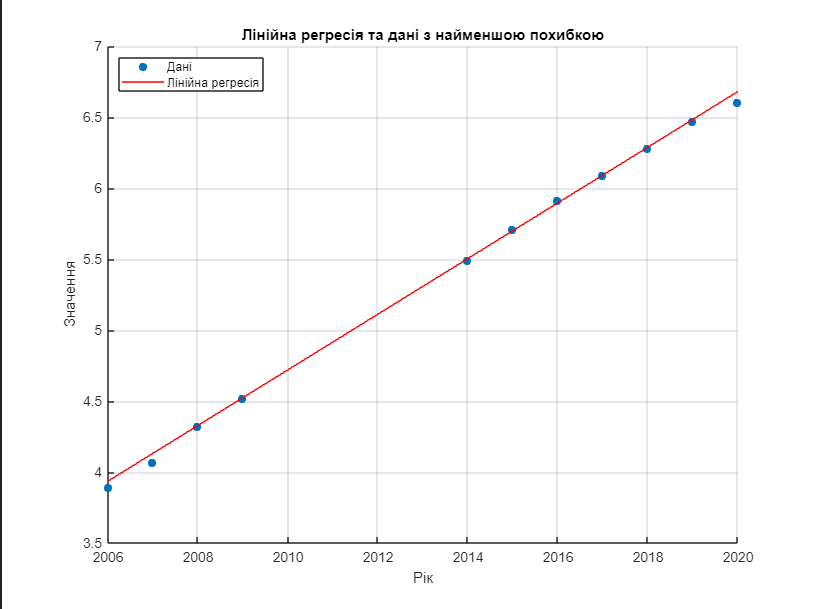
disp('Параметри лінійної регресії з найменшою похибкою:');

disp(['Коефіцієнт нахилу (a): ', num2str(best\_slope)]);

disp(['Зсув (b): ', num2str(best\_intercept)]);

disp(['Сумарна похибка на тестовій вибірці: ', num2str(error\_test)]);





6. Представити похибки лінійних регресій для даних тестової вибірки (дані представити у табличці, сформованій, аналогічно п.4d):

% Розрахувати прогнозовані значення для тестової вибірки

predicted\_values\_test = best\_slope \* year\_test + best\_intercept;

% Розрахувати сумарну похибку для тестової вибірки

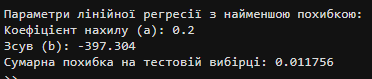
error\_test = sum((data\_values\_test - predicted\_values\_test).^2);

% Вивести результати аналізу регресії та похибки

disp('Результати аналізу регресії:');

results\_table = table(slope\_values', intercept\_values', error\_values', 'VariableNames', {'Коефіцієнт\_нахилу', 'Зсув', 'Сумарна\_похибка'});

disp(results\_table);



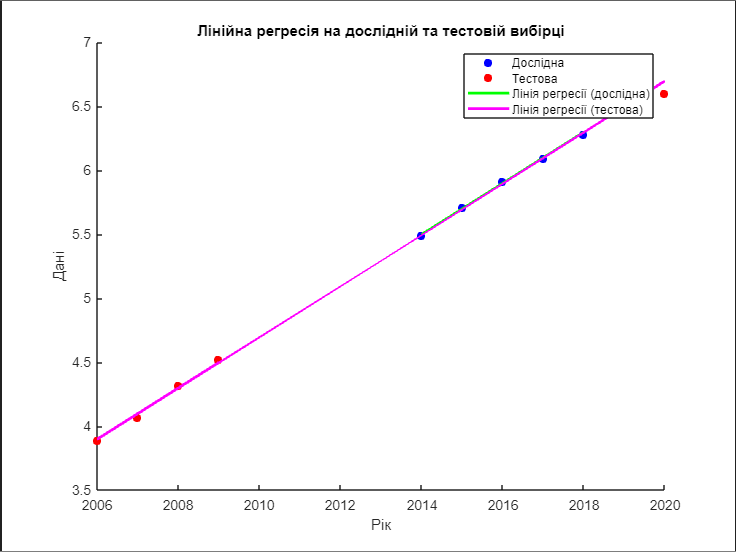
7. Побудувати графік, що відображає набір тестових даних, лінійну регресію:

xlabel('Рік');

ylabel('Дані');

title('Лінійна регресія на дослідній та тестовій вибірці');

legend('Дослідна', 'Тестова', 'Лінія регресії (дослідна)', 'Лінія регресії (тестова)');



Висновок: під час виконання завдання було проведено комплексний аналіз даних, що включав в себе ряд ключових етапів. Спочатку, дані були ретельно оброблені, і вони були розділені на дві частини: дослідну та тестову вибірки. Після цього, було визначено точку центру мас для кращого розуміння розподілу даних та їхніх характеристик. Найважливіше, були проведені дослідження лінійних регресій, що дозволило визначити взаємозв'язки та тенденції в даних, що можуть мати практичне застосування і важливе значення для подальших дій та рішень.