Міністерство освіти і науки України

Львівський національний університет імені Івана Франка

Факультет електроніки та комп’ютерних технологій

Кафедра системного проектування

Звіт

Про виконання лабораторної роботи №1

З курсу «Системи машинного навчання»

### Вступ в машинне навчання та Scikit-learn

**Виконала:**

Студентка групи ФЕС-32

Філь Дарина

**Перевірив:**

Доцент Колич І.І.

Львів 2024

**Мета:** Ознайомитись з базовими поняттями машинного навчання та бібліотекою Skilit-learn.

**Теоретичні відомості:**

Основні поняття:

1. **Залежна змінна (target, response)**: Це змінна, яку ми намагаємося передбачити або пояснити.
2. **Незалежні змінні (predictors, features):** Це змінні, які ми використовуємо для передбачення значення залежної змінної.

**Формула лінійної регресії Одновимірна лінійна регресія:**

Формула одновимірної (простої) лінійної регресії виглядає так:

де:

* — залежна змінна;
* — вільний член **(intercept);**
* — коефіцієнт нахилу **(slope);**
* — незалежна змінна;

**Множинна лінійна регресія:**

Формула множинної лінійної регресії (з більш ніж однією незалежною змінною) виглядає так:

де:

* — залежна змінна;
* — вільний член **(intercept);**
* — коефіцієнти регресії для кожної незалежної змінної;
* **—** незалежні змінні;

**Метод найменших квадратів (Ordinary Least Squares, OLS)**

Метод найменших квадратів (OLS) використовується для знаходження оптимальних значень коефіцієнтів регресії β̂, які мінімізують суму квадратів різниць між передбаченими значеннями та фактичними значеннями залежної змінної.

**Формула для обчислення коефіцієнтів регресії:**  
**де:**

* вектор оцінених коефіцієнтів;
* **𝑋** — матриця незалежних змінних;
* — вектор залежної змінної;
* **—** транспонована матриця ( X );
* **—** обернена матриця до ;

**Оцінка моделі**

Після навчання моделі лінійної регресії важливо оцінити її продуктивність. Ось деякі ключові метрики для оцінки моделі:

1. **Середньоквадратична помилка (Mean Squared Error, MSE):**

де:

* кількість спостережнь;

1. **Коефіцієнт детермінації :**

де:

показує, яка частка варіації залежної змінної може бути пояснена незалежними, змінними моделі. Значення варіюється від 0 до 1, де значення, близьке до 1, вказує на хорошу модель.

**Хід роботи:**

**Завдання**

1. Завантаження готових наборів даних з Scikit-learn:
   * Завантажити набір sklearn.datasets.fetch\_california\_housing
   * Виведіть перших 7 рядків, використовуючи бібліотеку Pandas
2. Поділ даних на тренувальну та тестову вибірки:

* Поділити дані на тренувальний 70% та тестовий 30% набори.

1. Написання функції для формули множинної лінійної регресії:

* Вихід залежна змінна (ціна будинку), вхід вектор незалежних змінних та вектор коефіцієнтів.

1. Випадковий підбір коефіцієнтів:

* Задати затравку (seed) для numpy використовуючи наступний код з заміною surname на прізвище автора.

my\_str = “surname”

res = “.join(format(ord(i), ‘08b’) for i in my\_str)

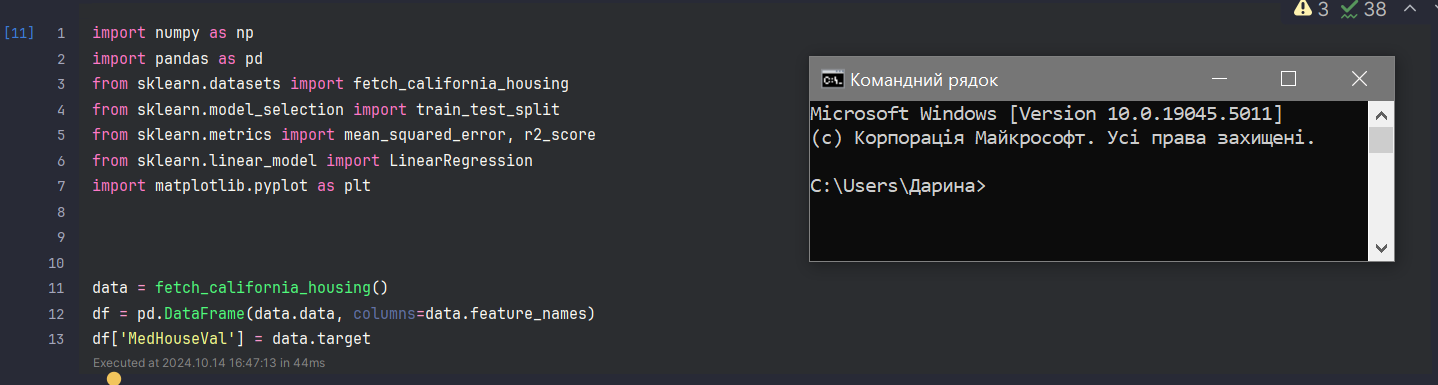
my\_seed = int(res) % 12345

* Написання цикл з генерацією випадкового вектору коефіцієнтів та вибрати вектор з найменшою середньоюквадратичною помилкою
* Оцінити коефіцієнт детермінації

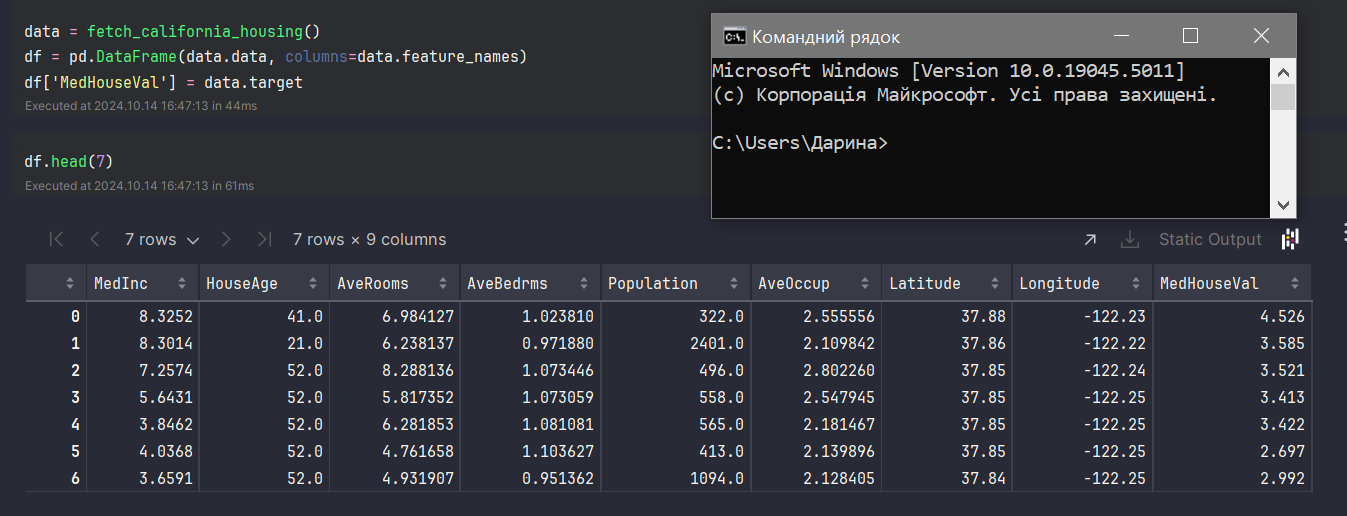
1. Навчання та оцінка простої моделі (наприклад, лінійна регресія):

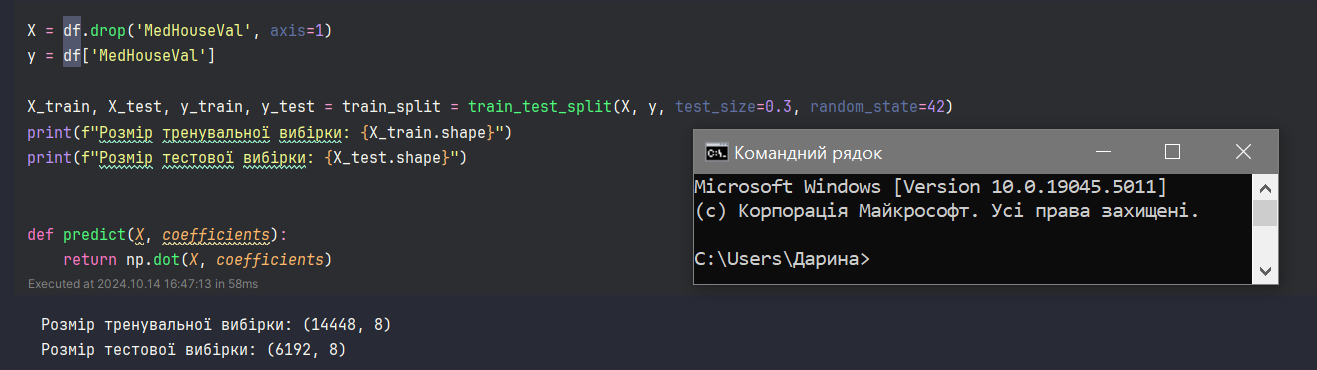
* Навчити моделі та тренувальних даних за допомогою лінійної регресії. Оцінити продуктивності моделі на тестових даних.

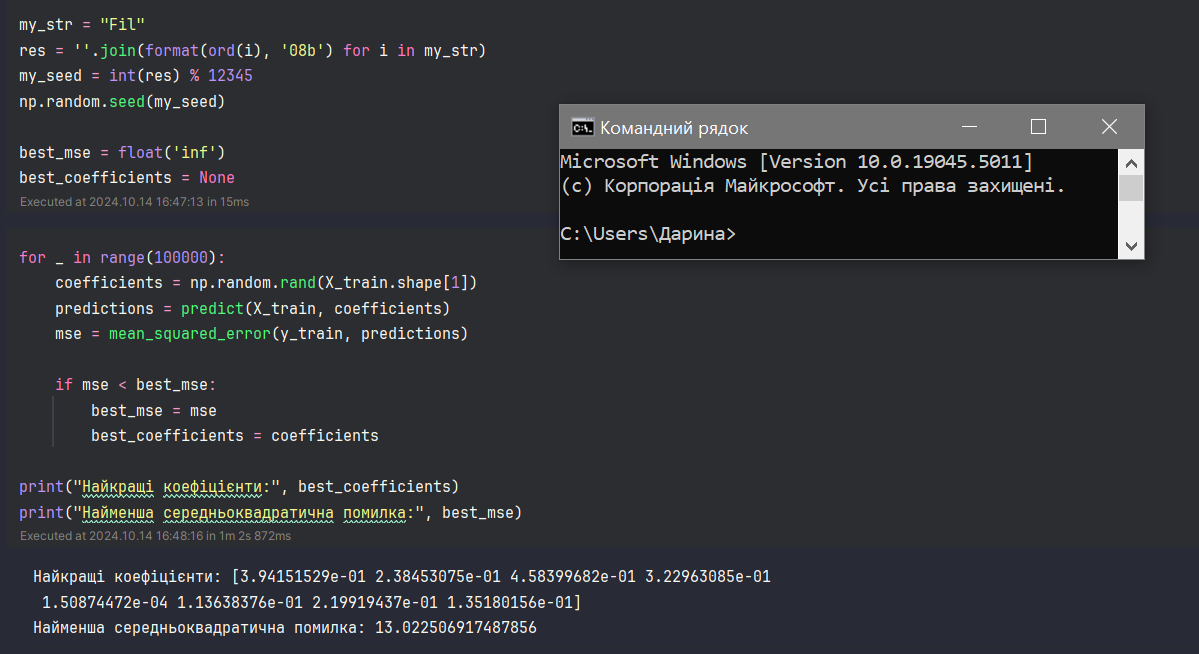
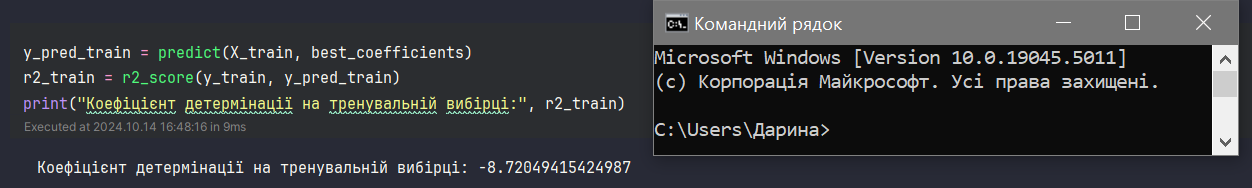
1. Оформити звіт.

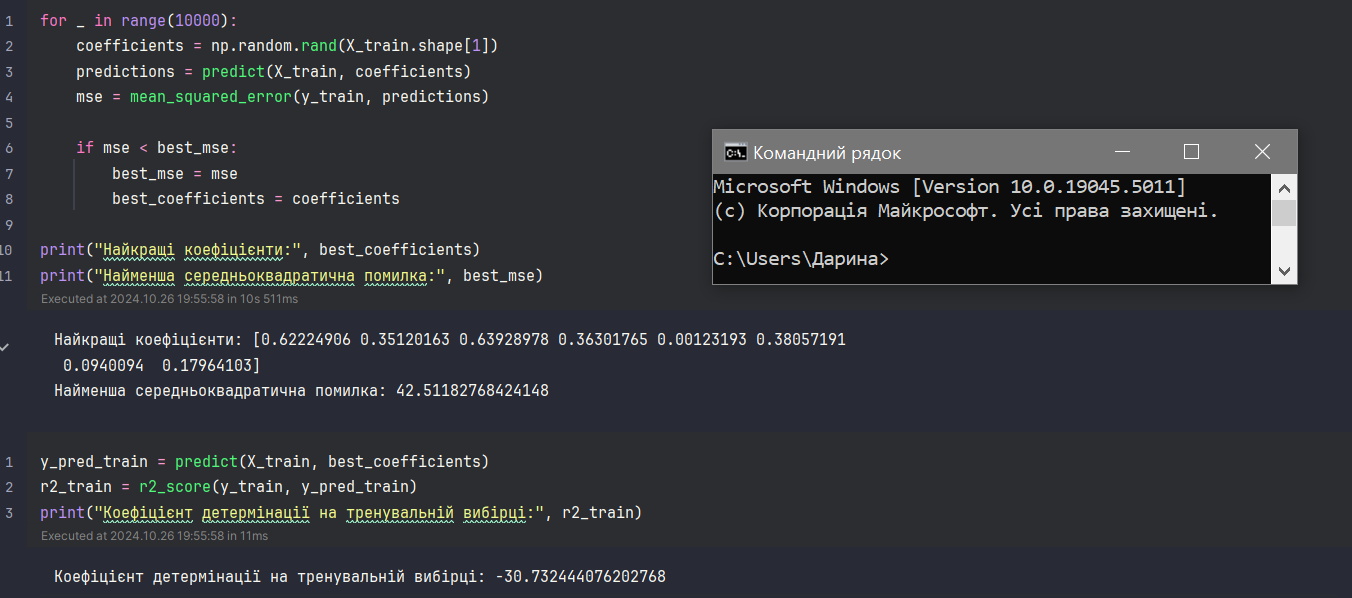


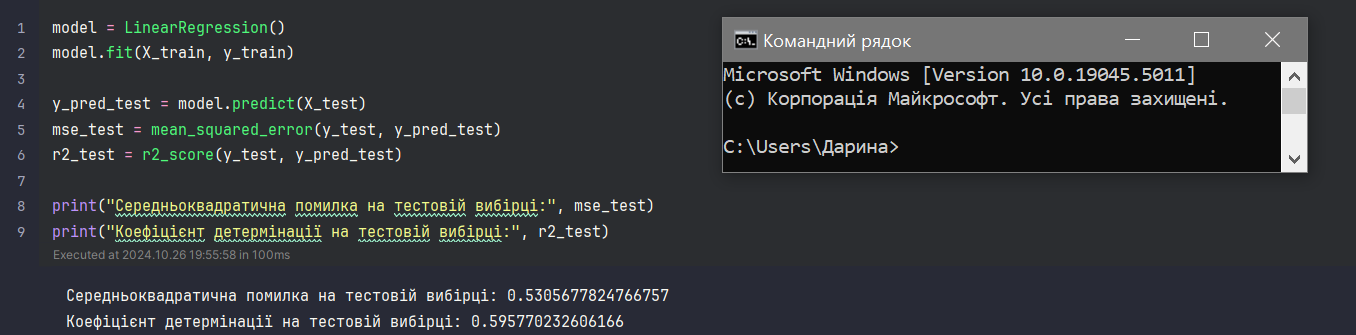
**Рис. 1** Ініціалізація усіх бібліотек та дата фрейму

  
  
**Рис. 2** Перші 7 значень дата фрейму

  
  
**Рис. 3** Поділ даних з дата фрейму на тестувальний та тренувальний. Також написана функція для множинної регресії

  
  
**Рис. 4** Створення seed та циклу для генерації випадкового вектору коефіцієнтів  
  
  
**Рис. 5** Тут зображено, що коефіцієнт детермінації на тренувальній виборці дуже поганий, це пов’язано з тим, що вибір коефіцієнтів є випадковим

  
  
**Рис. 6** При зменшенні розміру вектора можемо побачити, що помилка росте, а точність коефіцієнту детермінації падає ще більше

  
  
**Рис. 7** Результат навчання простої моделі лінійної регресії, можемо бачити, що коефіцієнт детермінації 0,59, що вказує на те, що більша частина варіації залежної змінної може бути пояснена незалежними змінними

**Висновок:** У цій лабораторній роботі я навчилась працювати з основами тренервування простих моделей та бібліотекою skilit-learn, також провела наглядний приклад тренерування моделі та оцінила результат тренерування на основі коефіцієнту детермінації та середній квадратичній похибці.