Міністерство освіти і науки україни

ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Математичний факультет

КАФЕДРА ПРОГРАМНОЇ ІНЖЕНЕРІЇ

Дисципліна «Емпіричні методи програмної інженерії»

Лабораторна робота № 4

Виконав: студент гр. 6.1219-2

Кабанов Костянтин

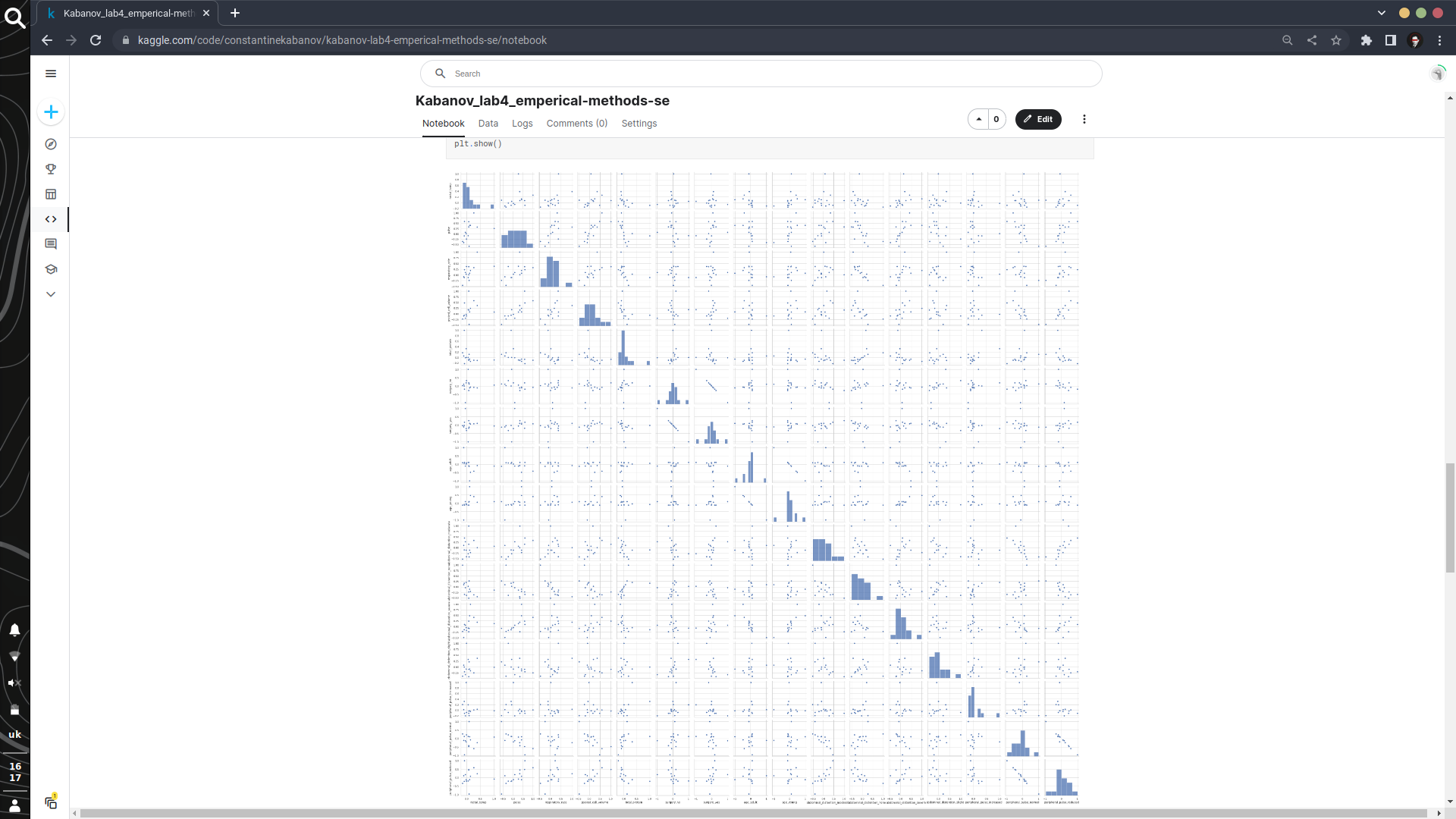
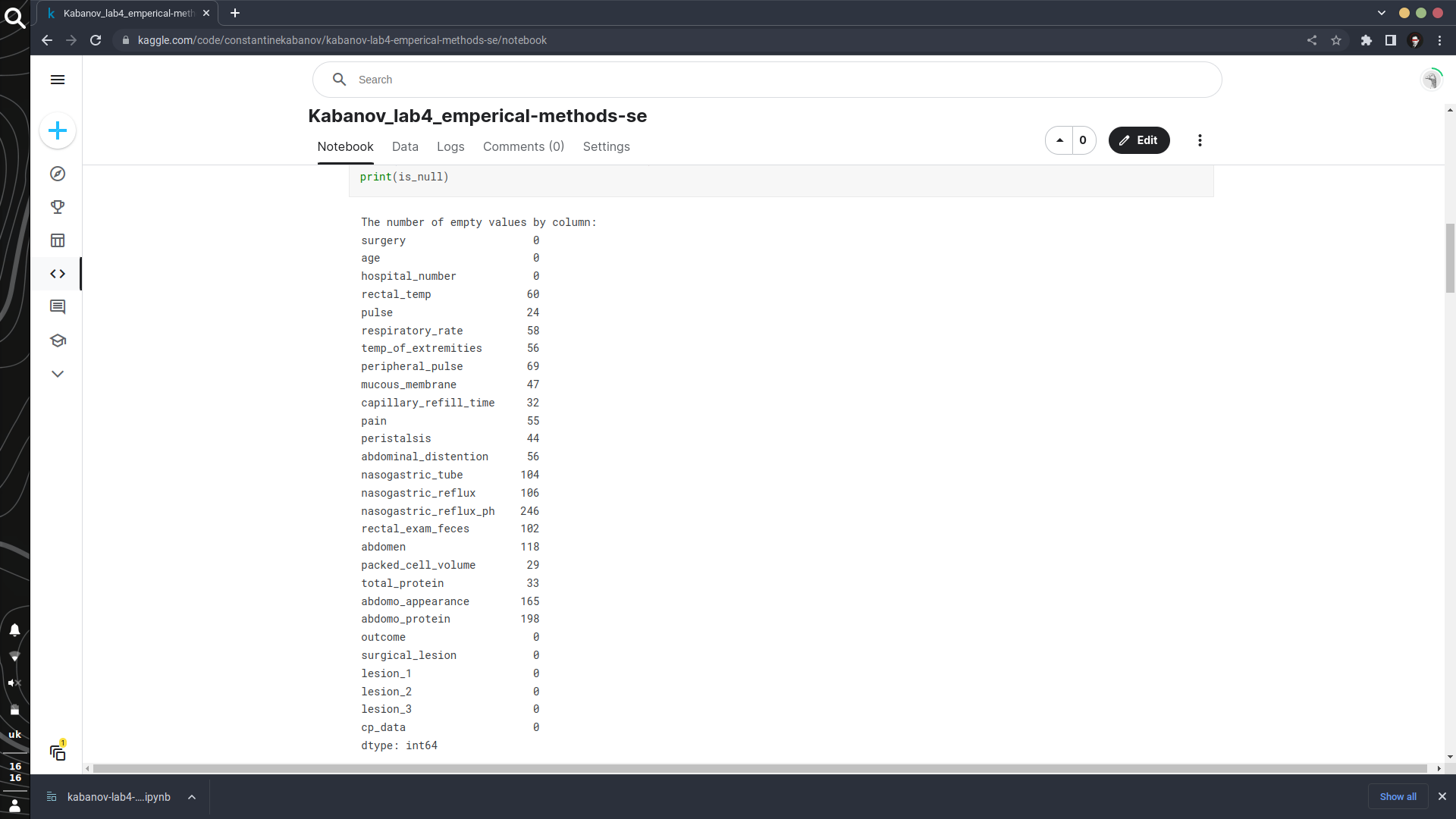
Перевірив: Викладач

Кривохата А.Г.

Запоріжжя

2022

**Мета:** засвоїти основні підходи до розв’язання задачі класифікації засобами Python.



**Посилання на код та розв’язок програми:**

**https://www.kaggle.com/code/constantinekabanov/kabanov-lab4-emperical-methods-se/notebook**

[**https://github.com/Constantijne/empirical-methods-se**](https://github.com/Constantijne/empirical-methods-se)

Контрольні запитання

1. Загальна постановка задачі класифікації.
2. Методи класифікації.
3. Які метрики використовуються при оцінці якості класифікації?

Відповіді

1. формалізована задача, яка містить множину об'єктів (ситуацій), поділених певним чином на класи. Задана скінченна множина об'єктів, для яких відомо, до яких класів вони належать. Ця множина називається вибіркою. До якого класу належать інші об'єкти невідомо. Необхідно побудувати такий алгоритм, який буде здатний класифікувати довільний об'єкт з вихідної множини.
   * Двокласова класифікація. Найпростіший в технічному відношенні випадок, який служить основою для вирішення складніших завдань.
   * Багатокласова класифікація. Коли число класів досягає багатьох тисяч (наприклад, при розпізнаванні ієрогліфів або злитого мовлення), завдання класифікації стає істотно важчим.
   * Непересічні класи.
   * Пересічні класи. Об'єкт може належати одночасно до декількох класів.
   * Нечіткі класи. Потрібно визначати ступінь належності об'єкта кожному з класів, звичайно це дійсне число від 0 до 1.
   * Характеристичний опис — найпоширеніший випадок. Кожен об'єкт описується набором своїх характеристик, які називаються ознаками. Ознаки можуть бути числовими або нечисловими.
   * Матриця відстаней між об'єктами. Кожен об'єкт описується відстанями до всіх інших об'єктів навчальної вибірки. З цим типом вхідних даних працюють деякі методи, зокрема, метод \* найближчих сусідів, метод потенційних функцій.
   * Часовий ряд або сигнал є послідовність вимірів у часі. Кожен вимір може представлятися числом, вектором, а в загальному випадку — характеристичним описом досліджуваного б'єкта в цей час часу.
   * Зображення або відеоряд.
   * Зустрічаються і складніші випадки, коли вхідні дані представляються у вигляді графів, текстів, результатів запитів до бази даних, і т. д. Як правило, вони приводяться до першого або другого випадку шляхом попередньої обробки даних та вилучення характеристик.