

## Лабораторна робота № 2

**Тема.** Кластеризація даних за допомогою методу  $k$ -середніх та класифікація даних за допомогою методу  $k$ -найближчих сусідів.

**Мета.** Ознайомитись з поняттям кластера, кластеризації; розібрати та реалізувати алгоритм  $k$ -середніх; ознайомитись з поняттям класифікації; розібрати та реалізувати алгоритм  $k$ -найближчих сусідів; навчитись застосовувати алгоритм для розв'язання задач класифікації.

### Виконання роботи

В ході роботи необхідно виконати наступні завдання:

1. Ознайомитись з алгоритмом  $k$ -середніх [1, с. 22-26].
2. Реалізувати алгоритм  $k$ -середніх у середовищі MatLab.
3. Реалізувати алгоритм  $k$ -середніх довільною мовою програмування.
4. Провести кластеризацію даних, утворених з часового ряду з ЛР 1, використовуючи оптимальну кількість кластерів (оцінити візуально). Дані побудувати за наступними правилами: (1) перетворити дані у набори точок  $(x_i, x_i+10)$ , (2) перетворити дані у набори точок  $(x_i, x_i+10, x_i+20)$ . Відобразити результати.
5. Розв'язати вправи 4, 5, 6, 8, 9 [1, с. 33-34].
6. Ознайомитись з алгоритмом  $k$ -найближчих сусідів [1, с. 152-155].
7. Реалізувати алгоритм  $k$ -найближчих сусідів у середовищі MatLab.
8. Реалізувати алгоритм  $k$ -найближчих сусідів довільною мовою програмування.
9. Розв'язати вправи 5, 10 [1, с. 158-159].

### Оцінка виконаних завдань

#### Вміння:

1. Реалізація програми  $k$ -середніх у середовищі MatLab (7%).
2. Побудова графіків вихідного набору даних та результату кластеризації з поясненням вибраної кількості кластерів та результату для завдання 4.1 (5%).
3. Побудова графіків вихідного набору даних та результату кластеризації з поясненням вибраної кількості кластерів та результату для завдання 4.2 (5%).
4. Реалізація програми  $k$ -найближчих сусідів у середовищі MatLab (6%).
5. Відповіді на вправи 4-6 (завдання 5) (3 x 5%).
6. Розв'язки вправ 8, 9 (завдання 5) (2 x 9%).
7. Відповіді на вправи 5, 10 (завдання 9) (2 x 9%).

«Аналіз даних»; ПМ4; 2016-2017 н.р.

8. Реалізація програми  $k$ -середніх довільною мовою програмування (13%).

9. Реалізація програми  $k$ -найближчих сусідів довільною мовою програмування (13%).

### **Звіт повинен містити**

1. Коди розроблених програм з коментарями.
2. Графіки з відповідними підписами по осям, легендою та поясненнями до них.
3. Теоретичні відомості, алгоритми, схеми, пояснення у випадку необхідності.

### **Список літератури для опрацювання**

1. Wu X. Kumar V. The top ten algorithms in data mining.2009.