

Лабораторна робота №5. Методи кластеризації

Підгрупа №1

1. Відкрити та зчитати наданий файл з даними.
2. Визначити та вивести кількість записів.
3. Видалити атрибут *quality*.
4. Вивести атрибути, що залишилися.
5. Використовуючи функцію **KMeans** бібліотеки **scikit-learn**, виконати розбиття набору даних на кластери з випадковою початковою ініціалізацією і вивести координати центрів кластерів.
Оптимальну кількість кластерів визначити на основі початкового набору даних трьома різними способами:
 - 1) elbow method;
 - 2) average silhouette method;
 - 3) prediction strength method (див. н. 9.2.3 *Determining the Number of Clusters* книжки **Andriy Burkov. The Hundred-Page Machine Learning Book**).Отримані результати порівняти і пояснити, який метод дав кращий результат і чому так (на Вашу думку).
6. За раніш обраної кількості кластерів багаторазово проведіть кластеризацію методом *k*-середніх, використовуючи для початкової ініціалізації метод *k*-means++.
Виберіть найкращий варіант кластеризації. Який кількісний критерій Ви обрали для відбору найкращої кластеризації?
7. Використовуючи функцію **AgglomerativeClustering** бібліотеки **scikit-learn**, виконати розбиття набору даних на кластери. Кількість кластерів обрати такою ж самою, як і в попередньому методі. Вивести координати центрів кластерів.
8. Порівняти результати двох використаних методів кластеризації.