Spring 2023

AIX20003 Project – K-Means Algorithm

Due date

Sunday, May 28th, 2023.

Please note that no late submissions will be accepted.

Project description

Given that the K-Means algorithm has been widely used in numerous applications, the objective of this project is to exercise implementing the K-Means algorithm from a simple toy problem to a complex real-world problem. The simulation code should be written in Python. The main deliverables for this project are 1) a presentation file outlining the storylines (15 minutes long) used to nail down this project and 2) simulation codes your team developed. All deliverables must be a joint effort, meaning that your team will need to distribute the different tasks amongst the members to share the workload fairly and effectively. However, please note that one student from your team should submit materials on behalf of the entire team.

Team

Students will work in a team of three or four students with the aim of developing teamwork skills while working with other students. The entire team will receive the same base grade; however, I will ask each of you to submit your own peer assessment in which you evaluate the contributions of your teammates; therefore, the grade may be changed as needed. This assessment aims to find ways to work well together and contribute equally to the overall product.

Data

Students will be required to use real-world operational datasets to complete this project. The datasets can be downloaded from the "Project" folder in the LMS system. Please note that no datasets are provided to address Task 1, but you will need to use the provided datasets for Task 2 and Task 3.

- ✈ South_Korea_territory.csv
- ✈ Vertiport_candidates.csv

Citation

Please ensure that all codes and materials originating from other sources including ChatGPT must be clearly documented.

Task 1: Toy problem

Assume that your team is given the following eight points (x,y) in the Cartesian coordinate system:

- Point 1 = (2,10)
- Point 2 = (2,5)
- Point 3 = (8,4)
- Point 4 = (5,8)
- Point 5 = (7,5)
- Point 6 = (6,4)
- Point 7 = (1,2)
- Point 8 = (4,9)

Your team is required to make three different clusters for the points using the K-Means algorithm. Suppose that three different centroid points are randomly pre-determined as follows:

- Centroid 1 = (2,10) is associated with Cluster 1
- Centroid 2 = (5,8) is associated with Cluster 2
- Centroid 3 = (1,2) is associated with Cluster 3

Your presentation should include the following items:

- ✈ How many iterations are needed to complete the clustering task?
- ✈ What is your team's strategy to stop the iterations?
- ✈ What do the clustering results look like?

Task 2: Open-ended problem

Assume that your team is tasked with evenly distributing N sample points onto the South Korea territory using the K-means algorithm. For example, you may need to equally distribute 10 sample points across the territory. Your presentation should include the following items:

- ✈ What is your team's strategy to complete the task?
- ✈ What does the resulting output look like?

Task 3: Real-world problem (i.e., Vertiport placement)

Suppose that your team is currently working for the Korean government. As the government has recently announced the first official roadmap for Regional Air Mobility (RAM) with the aim of introducing a new aviation transportation system, you are required to initiate a project to establish infrastructure for the early RAM operations. One of the key enablers for realizing RAM operations is to construct airfields for vehicles to take off and land, which is referred to as a vertiport. The construction of vertiport infrastructure requires careful consideration as building a vertiport is subject to land use, noise issues, or public safety.

(https://www.arbin.com/the-advantages-and-challenges-of-urban-air-mobility/ & https://transportup.com/tag/vertiport/)

Assume that initial vertiport locations are determined by aerospace professionals, which is given to your team with the file (i.e., Vertiport_candidates.csv). Your team's responsibility is to identify a proper number of vertiports as it is not possible to construct all the vertiport candidates at a time due to a limited budget. Your presentation should include the following items:

- ✈ Visualize the vertiport candidate locations in the Korean peninsula by using the given files (i.e., South_korea_territory.csv and Vertiport_candidates.csv)
- ✈ Let's say that the Korean government allows your team to assign only 17 vertiports in the Korean peninsula (i.e., K=17). Cluster the vertiport candidate locations using the K-Means algorithm and find the 17 centroid points, which become the final vertiport locations in this scenario.
- ✈ Imagine that the Korean government allows your team to place vertiports as many as your team wants, implying that the government does not have any financial issues. How many vertiports does your team want to establish in the Korean peninsula? Your team may not want to spend unnecessary building costs (e.g., consider all the given vertiports); but would like to invest money in an efficient manner (e.g., find an appropriate number of vertiports).