Spatial Analysis Using K-Means Clustering: From Toy Problems to Real-World Applications

# Introduction

In this document, we explore the application of K-Means clustering in spatial analysis. We will walk through three tasks that gradually increase in complexity and practicality, showcasing the versatility of K-Means clustering in various scenarios. The tasks covered are as follows:

1. Task 1: Toy Problem - Clustering Data Points

2. Task 2: Open-ended Problem - Distributing Sample Points

3. Task 3: Real-World Problem - Vertiport Placement Analysis

Task 1: Toy Problem - Clustering Data Points

This code is how to perform K-means clustering on a set of points. The points represent coordinates in the Cartesian coordinate system.

To start the clustering process, it is needed to specify the initial centroid points. These points act as the initial centers for the clusters. In this code, it is defined three initial centroid points. I created a K-Means object from the scikit-learn library, to perform the K-means clustering. I set the number of clusters to three since we have three initial centroid points. I also provide the initial centroid points to the algorithm.

We then fit the K-means model to the given points using the `fit` method. It performed the actual clustering and assigns each point to one of the clusters. After the clustering process is complete, we can access the cluster assignments for each point. These assignments indicate which cluster each point belongs to. In this code, we retrieve the cluster labels using the `labels\_` attribute of the fitted K-means model.

We can also obtain the coordinates of the cluster centers. These represent the average position of the points within each cluster. In this code, we retrieve the cluster centers using the `cluster\_centers\_` attribute of the fitted K-means model. Then I printed the cluster assignments for each point and the coordinates of the cluster centers, because we want to see the results of the clustering process.

In this task, we will start with a toy problem to understand the fundamental concepts of K-Means clustering. The steps involved are as follows:

1. Define the points: We begin by defining a set of data points in a two-dimensional space. These points could represent any dataset, such as customer locations or data measurements.

2. Define initial centroid points: Next, we select initial centroid points that will act as the starting positions for the clustering algorithm. These centroids should ideally be representative of the data distribution.

3. Perform K-Means clustering: Using the defined data points and initial centroids, we apply the K-Means algorithm. The algorithm iteratively assigns data points to the nearest centroid and updates the centroids until convergence is reached.

4. Retrieve cluster assignments and centers: After the K-Means algorithm converges, we obtain the cluster assignments for each data point, indicating which cluster it belongs to. Additionally, we retrieve the final coordinates of the cluster centers.

5. Visualize the results: To gain insights and assess the quality of the clustering, we can plot the data points with color-coded clusters and overlay the cluster centers on the plot. This visualization helps us understand the distribution of the data and the effectiveness of the clustering algorithm.

Task 2: Open-ended Problem - Distributing Sample Points

In this task, we move beyond toy problems and tackle an open-ended problem: distributing a given number of sample points across a region. The steps involved are as follows:

1. Load territory and sample point data: We load the necessary datasets, such as the territory information and the sample points to distribute. The territory data could include geographical coordinates or any other relevant information.

2. Extract relevant features: From the loaded data, we extract the necessary features required for the clustering analysis. These features could include geographical coordinates, population density, or any other attributes deemed important for the task.

3. Perform K-Means clustering: Using the extracted features, we apply the K-Means algorithm to distribute the sample points. The number of clusters is set to the desired number of sample points to be distributed.

4. Retrieve the coordinates of sample points: Once the clustering is complete, we obtain the coordinates of the resulting cluster centers. These coordinates represent the evenly distributed sample points across the region of interest.

5. Visualize the distribution: To visualize the distribution of the sample points, we can plot them on a map or any other appropriate visualization method. This helps in assessing the coverage and effectiveness of the clustering algorithm.

Task 3: Real-World Problem - Vertiport Placement Analysis

In Task 3, we address the real-world challenge of determining optimal locations for vertiports in a given region. Vertiports are crucial for air taxis and urban air mobility services. The objective is to strategically place vertiports to maximize accessibility and minimize travel distances, considering factors like population distribution and transportation infrastructure.

The steps involved in this task are:

1. Gather relevant data: Collect data on territory, population distribution, transportation infrastructure, regulations, and other factors influencing vertiport placement.

2. Preprocess the data: Clean, normalize, and extract features from the data, such as geographical coordinates and population density.

3. Perform K-Means clustering: Apply K-Means clustering to identify clusters of potential vertiport locations based on similar attributes.

4. Analyze cluster results: Evaluate cluster characteristics, considering factors like population density and proximity to transportation hubs.

5. Determine optimal vertiport locations: Select the best locations based on cluster analysis, population density, and travel patterns.

6. Validate and refine results: Validate findings using feasibility studies, safety regulations, expert feedback, and refine the results iteratively.

7. Visualize and communicate findings: Create visualizations and reports to effectively communicate the optimal vertiport locations and their benefits to stakeholders.

By leveraging K-Means clustering and considering various factors, we can make informed decisions on vertiport placement, contributing to efficient and accessible urban air mobility networks.