MSBA 511: Data Mining for Business Analytics

Conceptual Homework 2: Cluster Analysis

Objective

Be able to:

- Apply the steps of the k-Means clustering algorithm to identify two segments based on sample customer data.

Overview

A company wants to segment its customers based on their annual income and spending score on a scale of 1-100. For the purpose of this assignment, you do **not** need to normalize the data. The dataset for five customers is shown below:

Customer	Annual Income ($K)	Spending Score
A	15	39
B	16	81
C	28	15
D	29	40
E	33	70

.

Tasks

1. Use customers **A** and **D** as initial means for the **two** clusters. Perform as many iterations as necessary to assign each customer to the nearest cluster.

Submission Requirements

Submit a PDF document to Gradescope that includes the following:

1. A table showing each cluster and customers assigned to the cluster.
2. Describe each step, and show all relevant calculations. You may include screenshots or graphics in addition to a textual summary.
3. If customers **B** and **D** were used as initial means for the two clusters, describe how that would impact the steps for the k-Means algorithm and if it would change the cluster assignments for each customer.