Nanyang Technological University of Singapore

SC5002 Fundamentals of AI and Applications

Lab 3: Application of K-Means

Scenario

A retail company wants to segment its customers based on **Annual Income (k$)** and **Spending Score (1–100)**. The goal is to identify different customer groups for marketing campaigns.

The dataset (dummy data) represents **20 customers**:

Customer	Annual Income (X)	Spending Score (Y)
C1	15	39
C2	16	81
C3	17	6
C4	18	77
C5	19	40
C6	20	76
C7	21	6
C8	22	94
C9	23	15
C10	24	73
C11	25	88
C12	26	17
C13	27	40
C14	28	55
C15	29	82
C16	30	20

C17	31	79
C18	32	35
C19	33	83
C20	34	30

Tasks:

1. **Apply K-Means with k=3 and k=5**
    - Run the clustering algorithm on the dataset.
    - Visualize the clusters with different colors in a 2D scatterplot (X=Income, Y=Score).
    - Compare results between k=3 and k=5.
2. **Use the Elbow Method**
    - Compute **WCSS (within-cluster sum of squares)** for k=1,2,3,4,5,6.
    - Plot the **elbow curve** and determine the "optimal" number of clusters.
3. **Analyze Sensitivity to Initialization**
    - Run K-Means with the same k but different random seeds.
    - Compare cluster assignments and WCSS.
    - Discuss why results may vary.
4. **Visualize Clustering Limitations**

For the second synthetic dataset below with **two interlocking circles** (non-spherical clusters)

- Plot the scatter plot of this 30-point dataset (with true labels colored in blue and red; True label of 1: red; and True label of 0: blue)
- Apply K-Means with k=2.
- Visualize why K-Means fails in this case.
- Discuss why spherical assumption makes K-Means unsuitable for complex-shaped clusters.

Point	Feature1	Feature2	True Label

1	0.1331	-0.5126	1
2	0.4648	-0.1070	1
3	-0.5205	-0.0741	1
4	0.4625	0.3913	1
5	-0.9720	-0.2337	0
6	-0.8390	-0.5404	0
7	0.3491	-0.4034	1
8	-0.4556	-0.3020	1
9	0.2823	-0.9513	0
10	0.9885	0.0195	0
11	-0.0522	-0.9507	0
12	-0.4315	0.4549	1
13	0.1283	0.5358	1
14	-0.0563	-0.4657	1
15	0.4846	-0.0224	1
16	-0.1881	-0.9745	0
17	0.3259	0.9493	0
18	-0.8930	0.4404	0
19	-0.7778	-0.6721	0
20	0.6570	-0.7580	0
21	-0.0618	0.4373	1
22	0.6709	0.7405	0
23	-0.6336	-0.7034	0
24	-0.2857	-0.3625	1
25	-0.1296	0.4794	1
26	0.2557	-0.5007	1
27	-0.6574	0.7107	0

28	0.8932	-0.4518	0
29	-0.3771	0.2763	1
30	0.6122	-0.3904	1