

BÀI TẬP VỀ PHÂN CỤM

Câu 1. K-Means

- a) Trình bày ngắn gọn thuật toán k-means.
- b) Cho tập điểm: $x_1 = \{1, 3\}$, $x_2 = \{1.5, 3.2\}$, $x_3 = \{1.3, 2.8\}$, $x_4 = \{3, 1\}$, Dùng k-means để gom cụm với $k = 2$.
- c) Một xe đón khách về bến xe Miền Đông của công ty Mai Linh muốn đón n khách hàng. Do thời gian đón khách ít nên công ty muốn gom khách về k điểm để tiện việc đón. Giả sử $n = 5$ và $k = 2$. Năm khách hàng đang ở các tọa độ: A(1,1), B(3,1), C(3,3), D(4,2), E(1,3). Anh/chị hãy cho biết nên nhóm khách nào tới điểm đón nào để việc đưa đón là thuận tiện nhất. Cho biết tọa độ của 2 điểm cần đón khách? Giả sử độ đo khoảng cách được sử dụng là khoảng cách Euclidean.

Câu 2. Cho bảng dữ liệu sau:

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

- a) Áp dụng thuật toán k-means để phân cụm với $k=2$
- b) Áp dụng thuật toán Agglomerative Hierarchical Clustering để phân cụm dùng Complete Link, Single Link.

---o0o---

(Hết)