



Khai thác dữ liệu và ứng dụng

Bài tập nộp
GOM NHÓM
Mã bài tập: A4

1. Thuật toán AGNES

Cho tập dữ liệu gồm 8 điểm trong không gian 2 chiều : $P_1(2, 10)$, $P_2(2, 5)$, $P_3(8, 4)$, $P_4(5, 8)$, $P_5(7, 5)$, $P_6(6, 4)$, $P_7(1, 2)$, $P_8(4, 9)$:

- Hãy xây dựng ma trận khoảng cách giữa các điểm trên.
- Hãy sử dụng thuật toán **AGNES lần lượt** với **Single Link** và **Complete link** để gom nhóm (*trình bày chi tiết các bước*). Vẽ sơ đồ hình cây (dendogram) cho kết quả gom nhóm. (*Sơ đồ hình cây phải vẽ rõ ràng để nhận biết được thứ tự và giá trị của vị trí các nhóm gộp lại với nhau.*)
- Dựa trên sơ đồ hình cây tương ứng (theo Single Link và Complete Link) xác định **3 nhóm** thu được. So sánh kết quả.

2. Thuật toán K-means

Cho tập DL sau (đã chuẩn hoá) :

Customer	Age	Income	No. cards
Lâm	0.35	0.03	0.67
Hung	0.05	0.12	0.33
Mai	0.19	0.06	0.00
Lan	0.58	0.41	0.33
Thủy	0.00	0.00	0.67
Tuấn	0.33	0.15	0.33
Minh	1.00	1.00	0.00
Vân	0.81	0.65	0.33
Thiện	0.91	0.82	0.00
Ngọc	0.12	0.06	1.00

- Sử dụng thuật toán k-means (k=2) để gom nhóm. Trình bày rõ kết quả thu được.
- Tính độ đo SSE cho từng nhóm ở vòng lặp đầu tiên và cuối cùng.