

**Họ Tên:** Hà Huy Sơn

**MSSV:** 18574802010055

**Buổi thực hành 5: Phân cụm dữ liệu (K-Means)**

**Câu 1:** Cho tập dữ liệu gồm các mẫu có hai thuộc tính như sau:

S1[5.9, 3.2], S2[4.6, 2.9], S3[6.2, 2.8], S4[4.7, 3.2], S5[5.5, 4.2], S6[5.0, 3.0], S7[4.9 3.1], S8[6.7, 3.1], S9[5.1, 3.8], S10[6.0 3.0].

Phân cụm K-means với  $K = 3$  và độ đo khoảng cách giữa các điểm là khoảng cách Euclid. Các tâm cụm khởi tạo ban đầu C1(6.2,3.2); C2(6.6, 3.7); C3(6.5, 3.0).

Thực hiện các thao tác tính toán thủ công và trình bày kết quả tâm cụm sau mỗi lần lặp.

**Trả lời:**

Gọi  $C = \{C1, C2, C3\}$

- Ta có centroid:

+ C1(6.2, 3.2)

+ C2(6.6, 3.7)

+ C3(6.5, 3.0)

- **Lần lặp 1:**

+ S1[5.9, 3.2]

$$\text{dist}(S1, C1) = \sqrt{(6.2 - 5.9)^2 + (3.2 - 3.2)^2} = 0.3$$

$$\text{dist}(S1, C2) = \sqrt{(6.6 - 5.9)^2 + (3.7 - 3.2)^2} = 0.86$$

$$\text{dist}(S1, C3) = \sqrt{(6.5 - 5.9)^2 + (3.0 - 3.2)^2} = 0.632$$

$$\Rightarrow S1 \in C1$$

+ S2[4.6, 2.9]

$$\text{dist}(S2, C1) = \sqrt{(6.2 - 4.6)^2 + (3.2 - 2.9)^2} = 1.627$$

$$\text{dist}(S2, C2) = \sqrt{(6.6 - 4.6)^2 + (3.7 - 2.9)^2} = 2.154$$

$$\text{dist}(S2, C3) = \sqrt{(6.5 - 4.6)^2 + (3.0 - 2.9)^2} = 1.902$$

$$\Rightarrow S2 \in C1$$

+ S3[6.2, 2.8]

$$\text{dist}(S3, C1) = \sqrt{(6.2 - 6.2)^2 + (3.2 - 2.8)^2} = 0.4$$

$$\text{dist}(S3, C2) = \sqrt{(6.6 - 6.2)^2 + (3.7 - 2.8)^2} = 0.984$$

$$\text{dist}(S3, C3) = \sqrt{(6.5 - 6.2)^2 + (3.0 - 2.8)^2} = 0.36$$

$$\Rightarrow S3 \in C3$$

$$+ S4[4.7, 3.2]$$

$$\text{dist}(S4, C1) = \sqrt{(6.2 - 4.7)^2 + (3.2 - 3.2)^2} = 1.5$$

$$\text{dist}(S4, C2) = \sqrt{(6.6 - 4.7)^2 + (3.7 - 3.2)^2} = 1.96$$

$$\text{dist}(S4, C3) = \sqrt{(6.5 - 4.7)^2 + (3.0 - 3.2)^2} = 1.81$$

$$\Rightarrow S4 \in C1$$

$$+ S5[5.5, 4.2]$$

$$\text{dist}(S5, C1) = \sqrt{(6.2 - 5.5)^2 + (3.2 - 4.2)^2} = 1.22$$

$$\text{dist}(S5, C2) = \sqrt{(6.6 - 5.5)^2 + (3.7 - 4.2)^2} = 1.208$$

$$\text{dist}(S5, C3) = \sqrt{(6.5 - 5.5)^2 + (3.0 - 4.2)^2} = 1.562$$

$$\Rightarrow S5 \in C2$$

$$+ S6[5.0, 3.0]$$

$$\text{dist}(S6, C1) = \sqrt{(6.2 - 5.0)^2 + (3.2 - 3.0)^2} = 1.216$$

$$\text{dist}(S6, C2) = \sqrt{(6.6 - 5.0)^2 + (3.7 - 3.0)^2} = 1.746$$

$$\text{dist}(S6, C3) = \sqrt{(6.5 - 5.0)^2 + (3.0 - 3.0)^2} = 1.5$$

$$\Rightarrow S6 \in C1$$

$$+ S7[4.9, 3.1]$$

$$\text{dist}(S7, C1) = \sqrt{(6.2 - 4.9)^2 + (3.2 - 3.1)^2} = 1.303$$

$$\text{dist}(S7, C2) = \sqrt{(6.6 - 4.9)^2 + (3.7 - 3.1)^2} = 1.802$$

$$\text{dist}(S7, C3) = \sqrt{(6.5 - 4.9)^2 + (3.0 - 3.1)^2} = 1.603$$

$$\Rightarrow S7 \in C1$$

$$+ S8[6.7, 3.1]$$

$$\text{dist}(S8, C1) = \sqrt{(6.2 - 6.7)^2 + (3.2 - 3.1)^2} = 0.51$$

$$\text{dist}(S8, C2) = \sqrt{(6.6 - 6.7)^2 + (3.7 - 3.1)^2} = 0.61$$

$$\text{dist}(S8, C3) = \sqrt{(6.5 - 6.7)^2 + (3.0 - 3.1)^2} = 0.223$$

$$\Rightarrow S8 \in C3$$

+ S9[5.1, 3.8]

$$\text{dist}(S9, C1) = \sqrt{(6.2 - 5.1)^2 + (3.2 - 3.8)^2} = 1.252$$

$$\text{dist}(S9, C2) = \sqrt{(6.6 - 5.1)^2 + (3.7 - 3.8)^2} = 1.503$$

$$\text{dist}(S9, C3) = \sqrt{(6.5 - 5.1)^2 + (3.0 - 3.8)^2} = 1.612$$

$$\Rightarrow S9 \in C1$$

+ S10[6.0, 3.0]

$$\text{dist}(S10, C1) = \sqrt{(6.2 - 6.0)^2 + (3.2 - 3.0)^2} = 0.282$$

$$\text{dist}(S10, C2) = \sqrt{(6.6 - 6.0)^2 + (3.7 - 3.0)^2} = 0.922$$

$$\text{dist}(S10, C3) = \sqrt{(6.5 - 6.0)^2 + (3.0 - 3.0)^2} = 0.5$$

$$\Rightarrow S10 \in C1$$

**Vậy:** Ta thu được 3 cụm:

+ C1 = {S1, S2, S4, S6, S7, S9, S10}

+ C2 = {S5}

+ C3 = {S3, S8}

Cập nhật lại trọng tâm cụm:

$$C1 = \left( \frac{5.9+4.6+4.7+5.0+4.9+5.1+6.0}{7}, \frac{3.2+2.9+3.2+3.0+3.1+3.8+3.0}{7} \right) = (5.17, 3.17)$$

$$C2 = (5.5, 4.2)$$

$$C3 = \left( \frac{6.2+6.7}{2}, \frac{2.8+3.1}{2} \right) = (6.45, 2.95)$$

**- Lần lặp 2:**

+ S1[5.9, 3.2]

$$\text{dist}(S1, C1) = \sqrt{(5.17 - 5.9)^2 + (3.17 - 3.2)^2} = 0.730$$

$$\text{dist}(S1, C2) = \sqrt{(5.5 - 5.9)^2 + (4.2 - 3.2)^2} = 1.077$$

$$\text{dist}(S1, C3) = \sqrt{(6.45 - 5.9)^2 + (2.95 - 3.2)^2} = 0.604$$

$$\Rightarrow S1 \in C3$$

+ S2[4.6, 2.9]

$$\text{dist}(S2, C1) = \sqrt{(5.17 - 4.6)^2 + (3.17 - 2.9)^2} = 0.63$$

$$\text{dist}(S2, C2) = \sqrt{(5.5 - 4.6)^2 + (4.2 - 2.9)^2} = 1.58$$

$$\text{dist}(S2, C3) = \sqrt{(6.45 - 4.6)^2 + (2.95 - 2.9)^2} = 1.85$$

$$\Rightarrow S2 \in C1$$

$$+ S3[6.2, 2.8]$$

$$\text{dist}(S3, C1) = \sqrt{(5.17 - 6.2)^2 + (3.17 - 2.8)^2} = 1.09$$

$$\text{dist}(S3, C2) = \sqrt{(5.5 - 6.2)^2 + (4.2 - 2.8)^2} = 1.56$$

$$\text{dist}(S3, C3) = \sqrt{(6.45 - 6.2)^2 + (2.95 - 2.8)^2} = 0.29$$

$$\Rightarrow S3 \in C3$$

$$+ S4[4.7, 3.2]$$

$$\text{dist}(S4, C1) = \sqrt{(5.17 - 4.7)^2 + (3.17 - 3.2)^2} = 0.47$$

$$\text{dist}(S4, C2) = \sqrt{(5.5 - 4.7)^2 + (4.2 - 3.2)^2} = 1.28$$

$$\text{dist}(S4, C3) = \sqrt{(6.45 - 4.7)^2 + (2.95 - 3.2)^2} = 1.76$$

$$\Rightarrow S4 \in C1$$

$$+ S5[5.5, 4.2]$$

$$\text{dist}(S5, C1) = \sqrt{(5.17 - 5.5)^2 + (3.17 - 4.2)^2} = 1.08$$

$$\text{dist}(S5, C2) = \sqrt{(5.5 - 5.5)^2 + (4.2 - 4.2)^2} = 0$$

$$\text{dist}(S5, C3) = \sqrt{(6.45 - 5.5)^2 + (2.95 - 4.2)^2} = 1.57$$

$$\Rightarrow S5 \in C2$$

$$+ S6[5.0, 3.0]$$

$$\text{dist}(S6, C1) = \sqrt{(5.17 - 5.0)^2 + (3.17 - 3.0)^2} = 0.24$$

$$\text{dist}(S6, C2) = \sqrt{(5.5 - 5.0)^2 + (4.2 - 3.0)^2} = 1.3$$

$$\text{dist}(S6, C3) = \sqrt{(6.45 - 5.0)^2 + (2.95 - 3.0)^2} = 1.45$$

$$\Rightarrow S6 \in C1$$

$$+ S7[4.9, 3.1]$$

$$\text{dist}(S7, C1) = \sqrt{(5.17 - 4.9)^2 + (3.17 - 3.1)^2} = 0.27$$

$$\text{dist}(S7, C2) = \sqrt{(5.5 - 4.9)^2 + (4.2 - 3.1)^2} = 1.25$$

$$\text{dist}(S7, C3) = \sqrt{(6.45 - 4.9)^2 + (2.95 - 3.1)^2} = 1.55$$

$$\Rightarrow S7 \in C1$$

$$+ S8[6.7, 3.1]$$

$$\text{dist}(S8, C1) = \sqrt{(5.17 - 6.7)^2 + (3.17 - 3.1)^2} = 1.53$$

$$\text{dist}(S8, C2) = \sqrt{(5.5 - 6.7)^2 + (4.2 - 3.1)^2} = 1.62$$

$$\text{dist}(S8, C3) = \sqrt{(6.45 - 6.7)^2 + (2.95 - 3.1)^2} = 0.29$$

$$\Rightarrow S8 \in C3$$

$$+ S9[5.1, 3.8]$$

$$\text{dist}(S9, C1) = \sqrt{(5.17 - 5.1)^2 + (3.17 - 3.8)^2} = 0.63$$

$$\text{dist}(S9, C2) = \sqrt{(5.5 - 5.1)^2 + (4.2 - 3.8)^2} = 0.56$$

$$\text{dist}(S9, C3) = \sqrt{(6.45 - 5.1)^2 + (2.95 - 3.8)^2} = 1.59$$

$$\Rightarrow S9 \in C2$$

$$+ S10[6.0, 3.0]$$

$$\text{dist}(S10, C1) = \sqrt{(5.17 - 6.0)^2 + (3.17 - 3.0)^2} = 0.84$$

$$\text{dist}(S10, C2) = \sqrt{(5.5 - 6.0)^2 + (4.2 - 3.0)^2} = 1.3$$

$$\text{dist}(S10, C3) = \sqrt{(6.45 - 6.0)^2 + (2.95 - 3.0)^2} = 0.45$$

$$\Rightarrow S10 \in C3$$

**Vậy:** Sau bước lặp thứ 2 ta thu được 3 cụm:

$$+ C1 = \{S2, S4, S6, S7\}$$

$$+ C2 = \{S5, S9\}$$

$$+ C3 = \{S1, S3, S8, S10\}$$

Cập nhật lại trọng tâm cụm:

$$C1 = \left( \frac{4.6+4.7+5.0+4.9}{4}, \frac{2.9+3.2+3.0+3.1}{4} \right) = (4.8, 3.05)$$

$$C2 = \left( \frac{5.5+5.1}{2}, \frac{4.2+3.8}{2} \right) = (5.3, 4)$$

$$C3 = \left( \frac{5.9+6.2+6.7+6.0}{4}, \frac{3.2+2.8+3.1+3.0}{4} \right) = (6.2, 3.025)$$

**- Lần lặp 3:**

$$+ S1[5.9, 3.2]$$

$$\text{dist}(S1,C1) = \sqrt{(4.8 - 5.9)^2 + (3.05 - 3.2)^2} = 1.11$$

$$\text{dist}(S1,C2) = \sqrt{(5.3 - 5.9)^2 + (4 - 3.2)^2} = 1$$

$$\text{dist}(S1,C3) = \sqrt{(6.2 - 5.9)^2 + (3.025 - 3.2)^2} = 0.35$$

$$\Rightarrow S1 \in C3$$

$$+ S2[4.6, 2.9]$$

$$\text{dist}(S2,C1) = \sqrt{(4.8 - 4.6)^2 + (3.05 - 2.9)^2} = 0.25$$

$$\text{dist}(S2,C2) = \sqrt{(5.3 - 4.6)^2 + (4 - 2.9)^2} = 1.30$$

$$\text{dist}(S2,C3) = \sqrt{(6.2 - 4.6)^2 + (3.025 - 2.9)^2} = 1.61$$

$$\Rightarrow S2 \in C1$$

$$+ S3[6.2, 2.8]$$

$$\text{dist}(S3,C1) = \sqrt{(4.8 - 6.2)^2 + (3.05 - 2.8)^2} = 1.422$$

$$\text{dist}(S3,C2) = \sqrt{(5.3 - 6.2)^2 + (4 - 2.8)^2} = 1.5$$

$$\text{dist}(S3,C3) = \sqrt{(6.2 - 6.2)^2 + (3.025 - 2.8)^2} = 0.225$$

$$\Rightarrow S3 \in C3$$

$$+ S4[4.7, 3.2]$$

$$\text{dist}(S4,C1) = \sqrt{(4.8 - 4.7)^2 + (3.05 - 3.2)^2} = 0.18$$

$$\text{dist}(S4,C2) = \sqrt{(5.3 - 4.7)^2 + (4 - 3.2)^2} = 1$$

$$\text{dist}(S4,C3) = \sqrt{(6.2 - 4.7)^2 + (3.025 - 3.2)^2} = 1.51$$

$$\Rightarrow S4 \in C1$$

$$+ S5[5.5, 4.2]$$

$$\text{dist}(S5,C1) = \sqrt{(4.8 - 5.5)^2 + (3.05 - 4.2)^2} = 1.34$$

$$\text{dist}(S5,C2) = \sqrt{(5.3 - 5.5)^2 + (4 - 4.2)^2} = 0.28$$

$$\text{dist}(S5,C3) = \sqrt{(6.2 - 5.5)^2 + (3.025 - 4.2)^2} = 1.36$$

$$\Rightarrow S5 \in C2$$

$$+ S6[5.0, 3.0]$$

$$\text{dist}(S6,C1) = \sqrt{(4.8 - 5.0)^2 + (3.05 - 3.0)^2} = 0.21$$

$$\text{dist}(S6,C2) = \sqrt{(5.3 - 5.0)^2 + (4 - 3.0)^2} = 1.044$$

$$\text{dist}(S6, C3) = \sqrt{(6.2 - 5.0)^2 + (3.025 - 3.0)^2} = 1.2$$

$$\Rightarrow S6 \in C1$$

$$+ S7[4.9 \ 3.1]$$

$$\text{dist}(S7, C1) = \sqrt{(4.8 - 4.9)^2 + (3.05 - 3.1)^2} = 0.111$$

$$\text{dist}(S7, C1) = \sqrt{(5.3 - 4.9)^2 + (4 - 3.1)^2} = 0.98$$

$$\text{dist}(S7, C1) = \sqrt{(6.2 - 4.9)^2 + (3.025 - 3.1)^2} = 1.302$$

$$\Rightarrow S7 \in C1$$

$$+ S8[6.7, 3.1]$$

$$\text{dist}(S8, C1) = \sqrt{(4.8 - 6.7)^2 + (3.05 - 3.1)^2} = 1.9$$

$$\text{dist}(S8, C2) = \sqrt{(5.3 - 6.7)^2 + (4 - 3.1)^2} = 1.66$$

$$\text{dist}(S8, C3) = \sqrt{(6.2 - 6.7)^2 + (3.025 - 3.1)^2} = 0.505$$

$$\Rightarrow S8 \in C3$$

$$+ S9[5.1, 3.8]$$

$$\text{dist}(S9, C1) = \sqrt{(4.8 - 5.1)^2 + (3.05 - 3.8)^2} = 0.807$$

$$\text{dist}(S9, C2) = \sqrt{(5.3 - 5.1)^2 + (4 - 3.8)^2} = 0.28$$

$$\text{dist}(S9, C3) = \sqrt{(6.2 - 5.1)^2 + (3.025 - 3.8)^2} = 1.345$$

$$\Rightarrow S9 \in C2$$

$$+ S10[6.0 \ 3.0].$$

$$\text{dist}(S10, C1) = \sqrt{(4.8 - 6.0)^2 + (3.05 - 3.0)^2} = 1.201$$

$$\text{dist}(S10, C1) = \sqrt{(5.3 - 6.0)^2 + (4 - 3.0)^2} = 1.22$$

$$\text{dist}(S10, C1) = \sqrt{(6.2 - 6.0)^2 + (3.025 - 3.0)^2} = 0.201$$

$$\Rightarrow S10 \in C3$$

Vậy: Sau bước lặp 3 ta thu được 3 cụm:

$$+ C1 = \{S2, S4, S6, S7\}$$

$$+ C2 = \{S5, S9\}$$

$$+ C3 = \{S1, S3, S8, S10\}$$

**Nhận xét:**

- Kết quả phân cụm giữ nguyên sau 3 lần lặp, giải thuật dừng và cho kết quả phân cụm:
  - $C1 = \{S2, S4, S6, S7\}$
  - $C2 = \{S5, S9\}$
  - $C3 = \{S1, S3, S8, S10\}$