

Bài tập về nhà

Buổi số: 09 + 10

Bài 1:

Dựa trên những kiến thức và tìm hiểu của em về bài toán phân cụm (clustering), em hãy lấy một ví dụ liên quan tới một vấn đề trong lĩnh vực IoT mà em có thể giải quyết thông qua việc sử dụng bài toán phân cụm. Em hãy giải thích về lý do em lựa chọn và nêu những ưu điểm của phân cụm khi em áp dụng để giải quyết vấn đề này.

Bài 2:

Phân cụm (clustering) có nhiều các ứng dụng khác nhau trong nhiều lĩnh vực. Em hãy lựa chọn và trình bày về một ứng dụng của phân cụm mà em quan tâm nhất.

Bài 3: (Giải bài toán bằng bút và máy tính cầm tay)

Để xây dựng một tòa nhà văn phòng thông minh, một công ty đã tiến hành thu thập được một tập dữ liệu từ những nhân viên làm việc trong văn phòng thông qua khảo sát (xem Bảng 1). Mặc dù vậy có nhiều nhân viên đã không điền đầy đủ thông tin trong mục Mức độ hài lòng.

Bảng 1: Tập dữ liệu thu được từ những nhân viên làm việc trong văn phòng.

Mẫu	Nhiệt độ (°C)	Độ ẩm (%)	Mức độ hài lòng
S1	30	60	
S2	25	70	Dễ chịu
S3	35	80	
S4	28	75	
S5	33	85	Khó chịu
S6	26	65	
S7	36	90	
S8	27	68	Dễ chịu
S9	32	55	Tuyệt vời

Em hãy sử dụng thuật toán k-means để tự động điền các thông tin còn thiếu vào cột Mức độ hài lòng trong Bảng 1. Giả sử rằng mức độ hài lòng của các nhân viên chỉ gồm có 3 mức là: Khó chịu, Dễ chịu, và Tuyệt vời.

Gợi ý: Mỗi mẫu dữ liệu gồm 2 đặc trưng (Nhiệt độ và Độ ẩm) và Mức độ hài lòng có thể được coi là nhãn (để sử dụng trong các bài toán học có giám sát khác).

Bài 3: (Thực hành với Python)

Phân đoạn hình ảnh (Image Segmentation) đề cập đến việc phân chia một hình ảnh thành nhiều phân đoạn (vùng) dựa trên sự giống nhau của các giá trị pixel. Phân đoạn hình ảnh thường được sử dụng để phát hiện đối tượng hoặc phân biệt các phần khác nhau của hình ảnh. Một bạn sinh viên đã áp dụng thuật toán phân cụm k-means với $k = 3$ để phân cụm bức ảnh **lenna.png** (Hình 1).

Dựa trên ví dụ của bạn sinh viên đã thực hiện (file **Image_Segmentation.ipynb**), em hãy thực hiện các tác vụ sau:

- Em hãy sửa đổi giá trị của số lượng cụm để xem số lượng cụm khác nhau ảnh hưởng đến việc phân đoạn hình ảnh như thế nào. (Hãy thử với số lượng cụm khác nhau, ví dụ 2, 4, 7, ...). Việc tăng số lượng cụm ảnh hưởng đến mức độ chi tiết trong hình ảnh được phân đoạn như thế nào?
- Theo mặc định, k-means sử dụng phương pháp khởi tạo k-means++. Thay đổi phương pháp khởi tạo thành ngẫu nhiên và quan sát sự khác biệt. Phương pháp khởi tạo ảnh hưởng như thế nào đến kết quả phân đoạn và tốc độ hội tụ?
- Sau khi thực hiện phân cụm k-means, hãy trực quan hóa các màu chủ đạo (trung tâm cụm) biểu diễn các vùng được phân đoạn. Các màu chủ đạo (trung tâm cụm) được xác định bởi k-means trong hình ảnh là gì?



Hình 1: Bức ảnh về Lenna (hay Lena) được sử dụng rộng rãi trong lĩnh vực xử lý ảnh¹.

¹ <https://en.wikipedia.org/wiki/Lenna> Bên cạnh đó, sinh viên nào quan tâm nhiều hơn về bức ảnh có thể tham khảo thêm thông tin qua Internet.