**Bài 02 Buổi 04 Bài tập thêm**

Có nhiều thuật toán gom cụm tự động xác định được số lượng cụm mà không cần

người dùng phải cho trước như k-Means. Nhiều người cho rằng đây là điểm vượt

trội của chúng so với k-Means. Em hãy liệt kê ít nhất 2 trường hợp để phản biện lại

quan điểm này.

* Chúng ta không thể quyết định số cụm mà chúng ta mong muốn, số lượng cụm mà chúng ta muốn phân loại cũng là một thông số quan trọng trong tập dataset. Việc không thể quyết định cụm cũng làm cho chúng ta khó có thể tùy biến cho phù hợp với từng tình huống.
* Một số phương pháp gom cụm không cần xác định số lượng cụm như là: “hierarchical clustering”. Điểm mạnh là nó không cần phải xác định số cụm nhưng nó có điểm yếu là hiếm khi đưa ra được kết quả tốt nhất, không chạy được với dữ liệu thiếu hoặc dữ liệu hỗn hợp. Khi tập dữ liệu lớn thì tốc độ chạy không được tốt.