

Clustering

- **Clustering (phân cụm)** là một **thuật toán học không giám sát (unsupervised learning)** có khả năng tự động nhóm các điểm dữ liệu có đặc điểm tương tự lại với nhau, không cần nhãn (**label**).
- Mục đích là tìm ra các **cụm (clusters)**: Nhóm các điểm gần nhau, tương tự nhau.

Supervised Learning	Unsupervised Learning (Clustering)
Dữ liệu gồm (x, y) : có cả đặc trưng và nhãn	Dữ liệu chỉ có x , không có nhãn y
Mục tiêu: học hàm dự đoán y từ x	Mục tiêu: tìm ra cấu trúc/nhóm ẩn trong x
Ví dụ: Phân loại nhị phân bằng logistic regression	Ví dụ: Nhóm các điểm dữ liệu thành các cụm