

Học Máy Nâng Cao

Kiểm tra

- Đề án (nhóm): 50%
 - Mỗi nhóm 3 - 4 học viên sẽ báo cáo một chủ đề của môn học (nội dung báo cáo + demo hay ví dụ minh họa).
- Trắc nghiệm: 50%
 - 23 / 2/ 2023: thi giữa kỳ
 - 23/ 3/ 2023: thi cuối kỳ

Nội dung

Chương 1: GIỚI THIỆU MÁY HỌC

Chương 2: MÔ HÌNH CÂY QUYẾT ĐỊNH (Decision Tree)

Chương 3: MÔ HÌNH MẠNG NƠI RON (Neural Network)

Chương 4: MÔ HÌNH HỌC THEO NHÓM KẾT HỢP (Boosting / Adaboost)

Chương 5: MÔ HÌNH THUẬT GIẢI DI TRUYỀN ĐỐI VỚI BÀI TOÁN PHÂN LỚP (Genetic Algorithm)

Chương 6 : MÔ HÌNH HỌC TĂNG CƯỜNG (Reinforcement Learning)

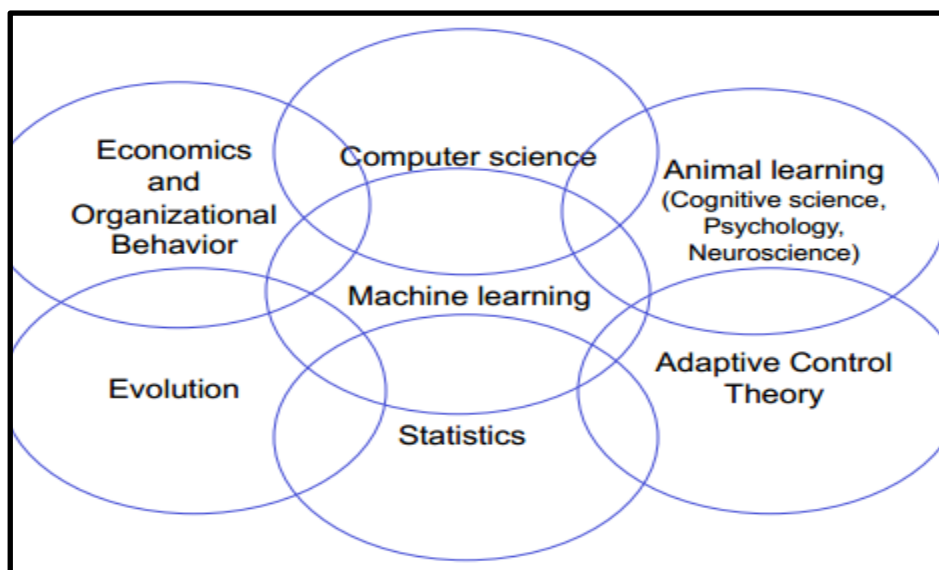
Chương 7: MÔ HÌNH HỌC MARKOV và MARKOV Ẩn

Chương 8: CÁC THUẬT TOÁN CLUSTERING

Giới thiệu Máy Học

MÁY HỌC

- Máy học dùng để chỉ đến việc **xác định, xây dựng và đánh giá** mô hình tính toán dùng để tái tạo, dự đoán hay ra quyết định từ tập dữ liệu cho trước



Tom M. Mitchell

MÁY HỌC

- Trước đây có khuynh hướng cho rằng việc nghiên cứu xây dựng và đánh giá các mô hình học hay thuật toán trong ML đóng vai trò quan trọng, nhưng gần đây có khuynh hướng cho rằng việc xác định làm thế nào sử dụng ML đóng vai trò quan trọng hơn trong nghiên cứu ML (“Machine Learning that matters” and “Trends in ML according to Google Scholars ”)

MÁY HỌC

Supervised
learning

Semi-Supervised
learning

Unsupervised
learning

Classification
Categorization

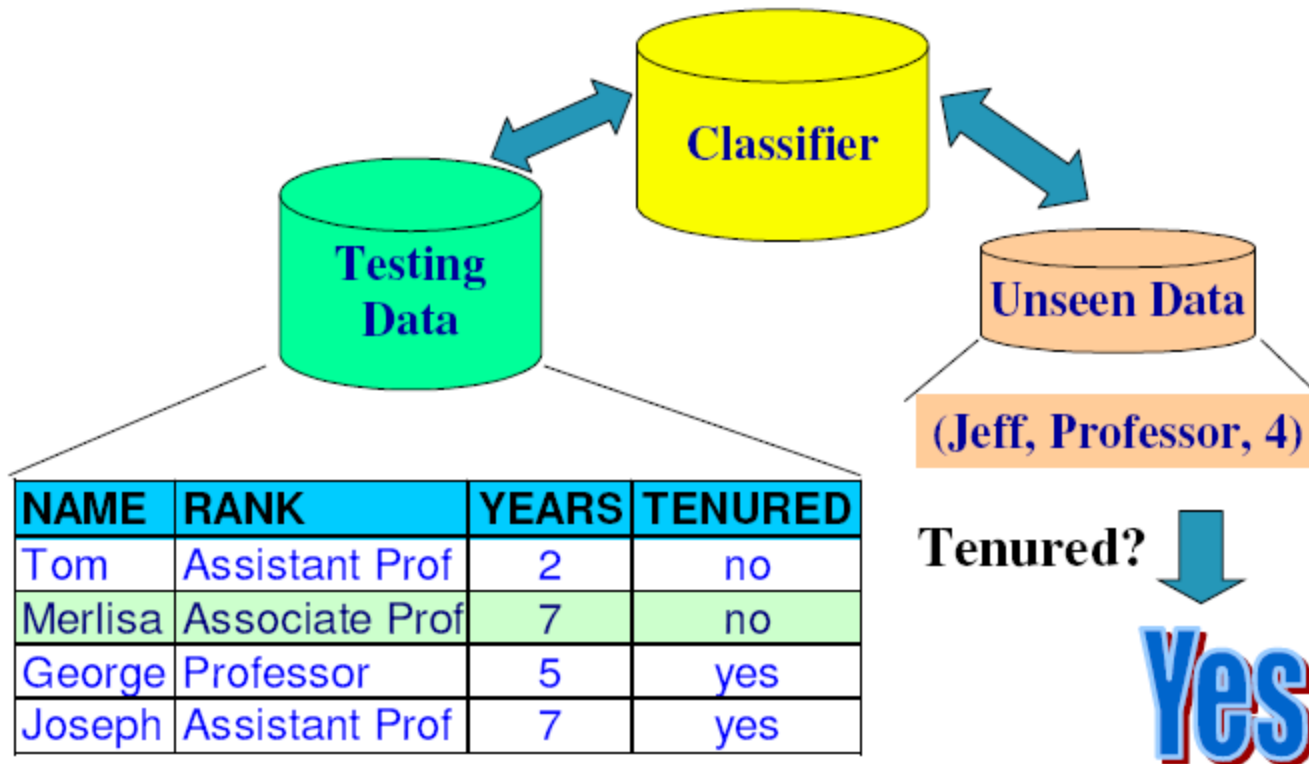


Clustering

Regression

**Dimensionality
Reduction**

MÔ HÌNH CÂY QUYẾT ĐỊNH



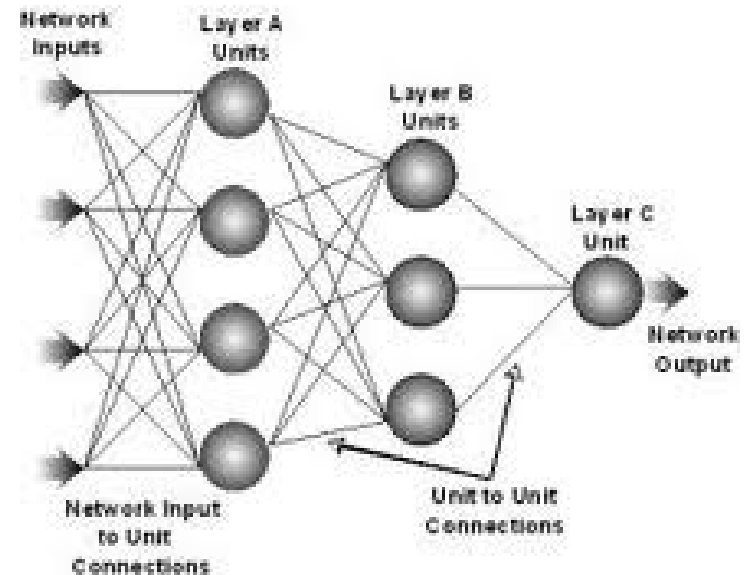
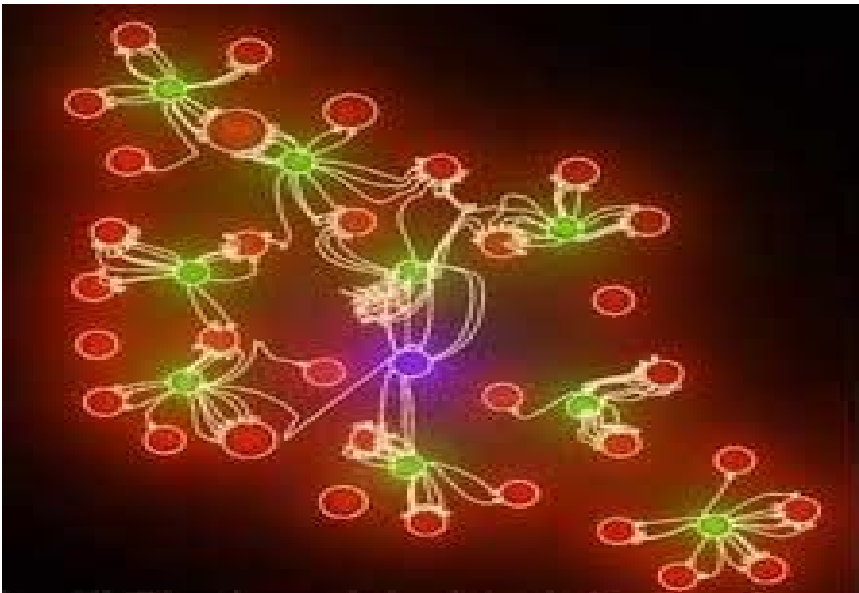
Cây quyết định là 1 phương pháp phân lớp thuộc nhóm học có giám sát, mục tiêu từ bảng dữ liệu cho trước xây dựng một cây quyết định từ tổ hợp giá trị các thuộc tính $X_1 X_2 \dots X_n$ có thể suy ra giá trị của thuộc tính Y

MÔ HÌNH CÂY QUYẾT ĐỊNH

- Các vấn đề trong việc xây dựng và áp dụng cây quyết định
 - Lựa chọn độ đo và xác định các node trong cây
 - Xử lý các data bị mất hay thiếu trong bảng dữ liệu
 - Xử lý trường hợp quá khớp (Overfitting)
 - Xử lý rút gọn cây qua quá trình sắp xếp lại và loại bỏ nhánh để tránh cây trùng lặp

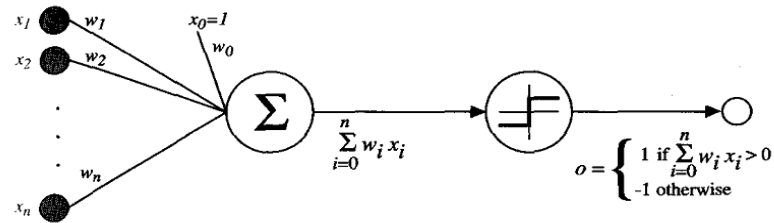
MÔ HÌNH MẠNG NƠI RON

- Mạng neural và ý nghĩa sinh học

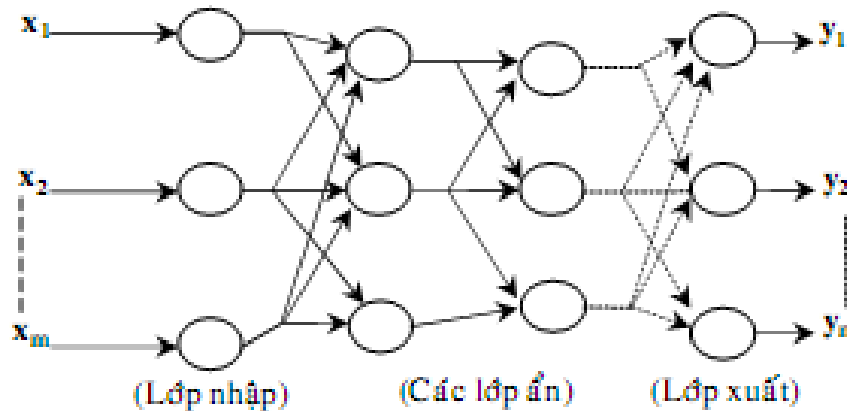


MÔ HÌNH MẠNG NƠ RON

- Mô hình perceptron

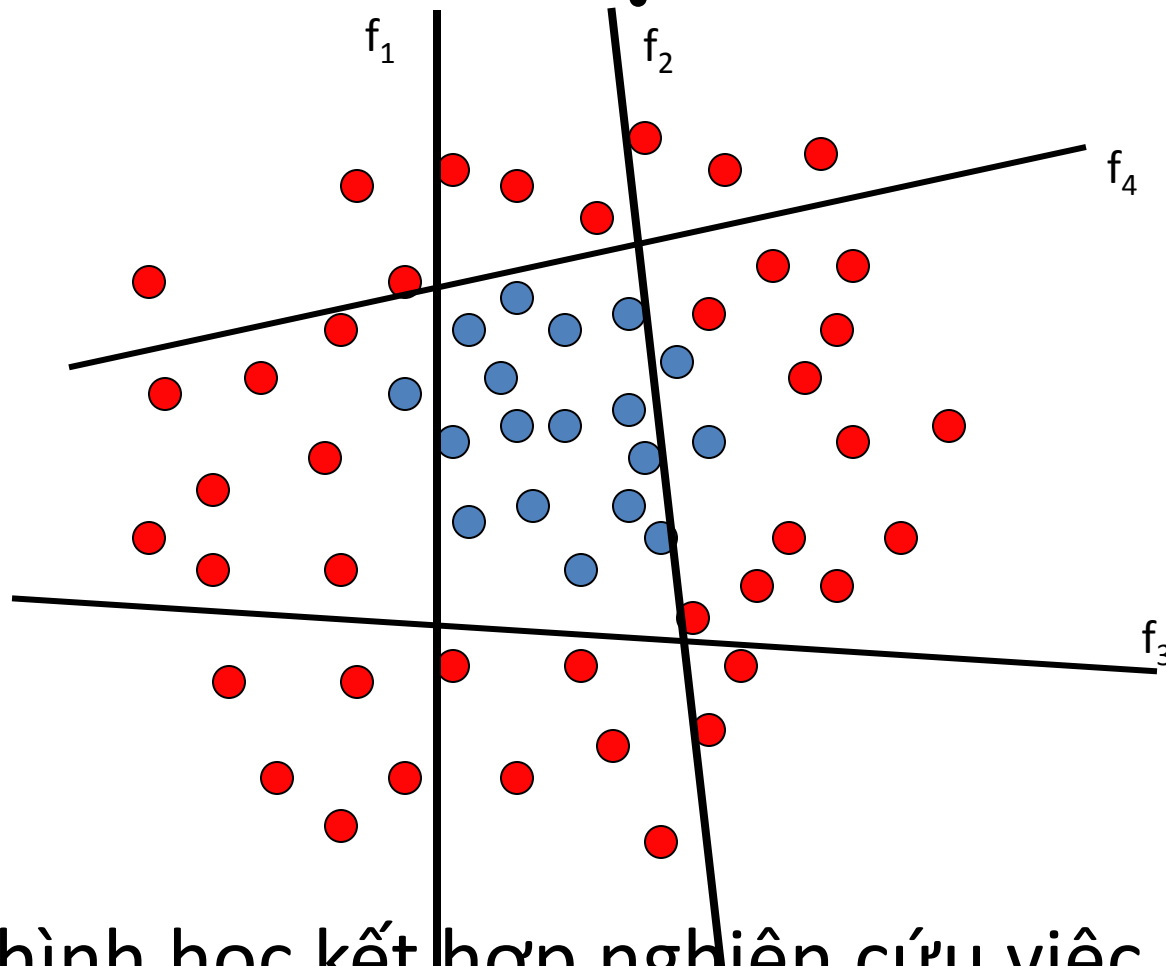


- Mô hình mạng đa lớp



- Mô hình mạng hai lớp
- Nghiên cứu các thuật toán học và đánh giá các mô hình trên

MÔ HÌNH HỌC THEO NHÓM KẾT HỢP



- Mô hình học kết hợp nghiên cứu việc kết hợp các hàm f_i để tạo thành bộ phân lớp tốt hơn

MÔ HÌNH HỌC THEO NHÓM KẾT HỢP

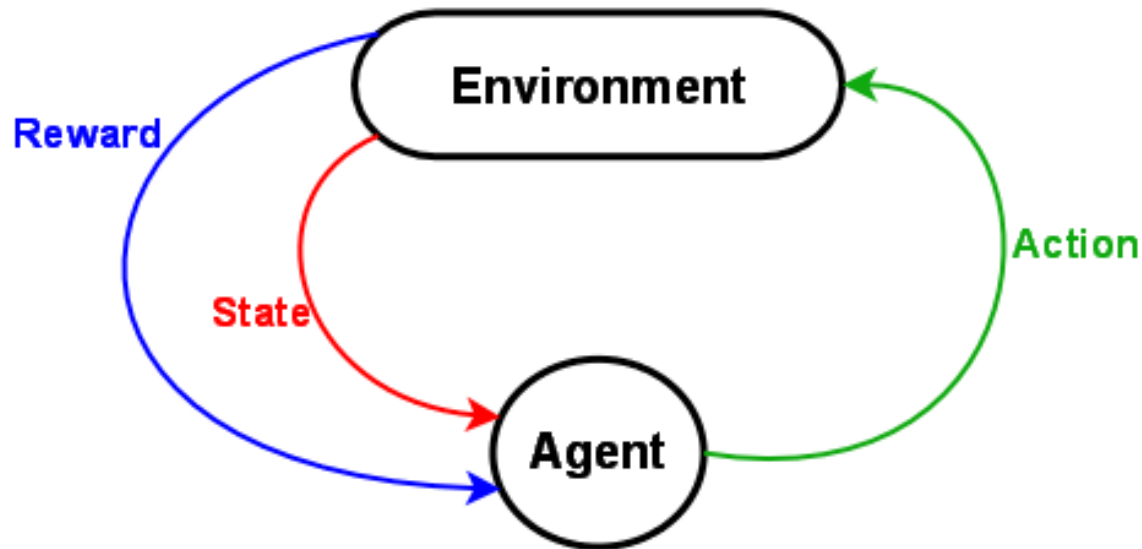
- Mô hình học mạnh : trong việc phân lớp đó là những phân lớp có độ chính xác cao mô hình yếu (không phải là 100%).
- Mô hình học yếu hơn: trong việc phân lớp đó là những phân lớp có độ chính xác không cao (nhưng không có nghĩa weak là quá thấp) phải hơn việc phân lớp ngẫu nhiên *một chút*.
- Vấn đề trong mô hình học theo nhóm kết hợp là xác định các trọng số kết hợp của các mô hình học yếu để tạo thành mô hình học mạnh hơn.
- Vấn đề phân chia bộ dữ liệu học để đạt được mô hình học mạnh hơn.

MÔ HÌNH THUẬT GIẢI DI TRUYỀN

- Xây dựng cấu trúc gen cho mỗi lời giải của bài toán, từ đó mã hoá thành một nhiễm sắc thể.
- Giải mã các nhiễm sắc thể thành lời giải của bài toán và đo lường mức độ lỗi của lời giải.
- Xây dựng thuật toán theo cơ chế di truyền để tìm kiếm các nhiễm sắc thể tốt trên không gian lời giải.

MÔ HÌNH HỌC TĂNG CƯỜNG

- Mô hình học tăng cường là mô hình xây dựng dựa trên mô hình sau



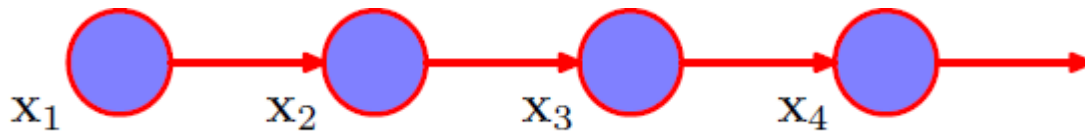
- Học tăng cường nghiên cứu cách thức một *agent* trong một môi trường nên chọn thực hiện các hành động nào để cực đại hóa điểm thưởng *reward*

MÔ HÌNH HỌC TĂNG CƯỜNG

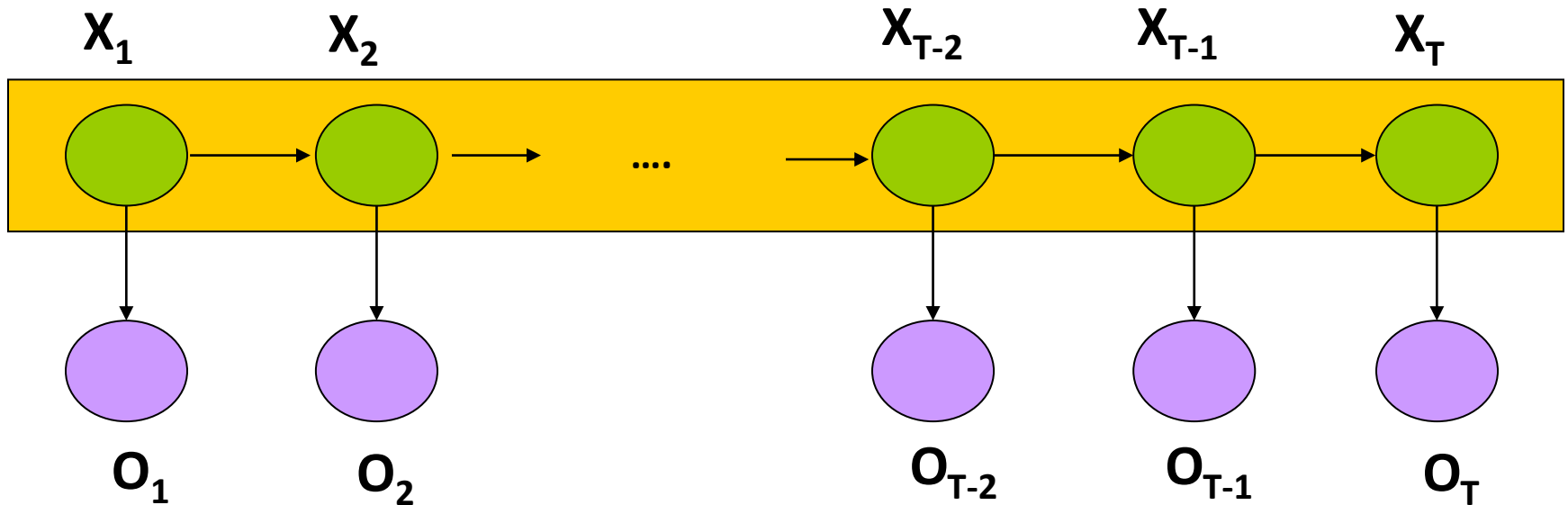
- Nghiên cứu các xác định môi trường, điểm thưởng và agent cho mô hình bài toán mà mình cần giải
- Thuật toán Q-learning và Sarsa

MÔ HÌNH MARKOV - MARKOV ẨN

- Mô hình Markov



- Mô hình Markov ẩn

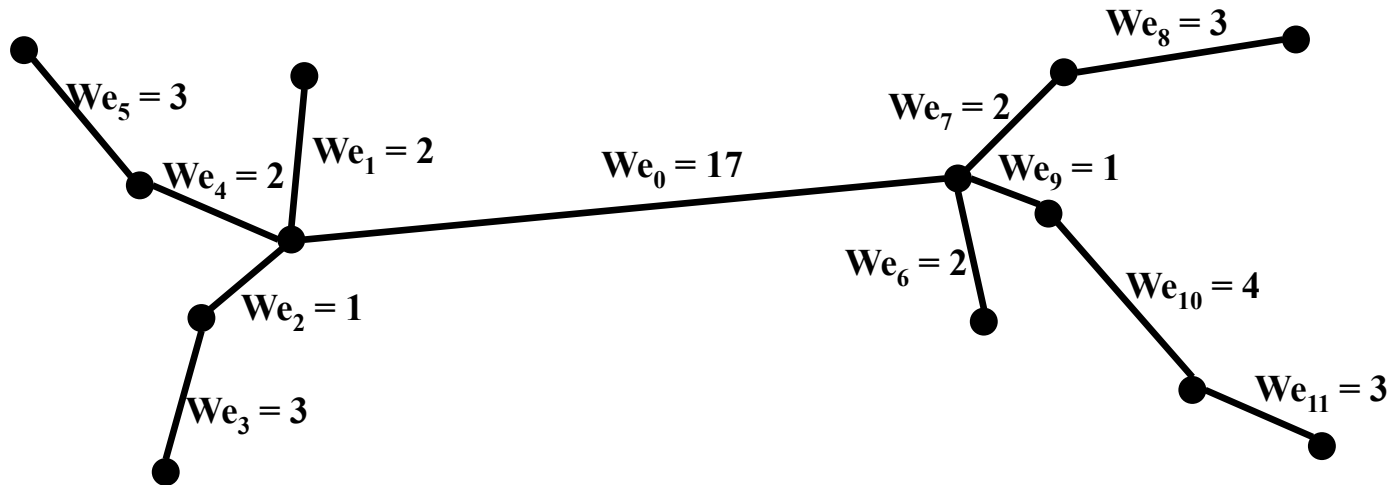


MÔ HÌNH MARKOV - MARKOV ẨN

- Xác định mô hình xác suất sử dụng trong mô hình Markov và Markov ẩn.
- Xác định các thuật toán học cho mô hình Markov ẩn và các thuật toán tính xác suất lân cận cực đại trong mô hình Markov và Markov ẩn.

CÁC THUẬT TOÁN CLUSTERING

- Các thuật toán phân nhóm dựa trên đồ thị



- Các thuật toán phân nhóm dựa trên vùng ảnh hưởng
- Các thuật toán branch and bound
-