

## BÀI TẬP VỀ PHÂN LỚP DỮ LIỆU

Nội dung:

- Phân lớp dựa trên Naïve Bayesian
- Phân lớp dựa trên cây quyết định
- Phân lớp dựa trên SVM
- Phân lớp dựa trên K-NN.

**Bài 1.** Cho cơ sở dữ liệu S:

	swims	has fins	flies	has lung	is a fish
Herring	yes	yes	no	no	yes
Cat	no	no	no	yes	no
Pigeon	no	no	yes	yes	no
Flying fish	yes	yes	yes	no	yes
Otter	yes	no	no	yes	no
Cod	yes	yes	no	no	yes
Whale	yes	yes	no	yes	no

a) Hãy xác định thuộc tính gốc cho cây quyết định đối với dữ liệu trên với thuộc tính phân lớp là “is a fish”?

b) Hãy vẽ cây quyết định cho tập dữ liệu trên.

c) Và giả sử cho thông tin về một loài động vật như sau:

$X = (\text{swims} = \text{yes}, \text{has fins} = \text{no}, \text{flies} = \text{yes}, \text{has lung} = \text{yes})$

Hãy nêu và áp dụng giải pháp phân lớp Naïve Bayesian để xác định X thuộc lớp “fish” hay “non-fish”?

**Bài 2.** Cho tập huấn luyện sau:

Số tt	Luyện tập	Ý chí	Năng khiếu	Đạt giải
1	Chăm	Có	Có	Có
2	Vừa phải	Có	Có	Có
3	Vừa phải	Không	Không	Không
4	Vừa phải	Không	Vừa phải	Không
5	Lười	Không	Có	Không
6	Chăm	Vừa phải	Có	Có
7	Chăm	Có	Vừa phải	Có
8	Lười	Vừa phải	Không	Không

Hãy vẽ cây quyết định cho tập huấn luyện trên?

### Bài 3. Luật Phân Lớp

Dùng thuật toán ID3 và Naïve Bayes để tìm luật phân lớp trong bảng sau đây.

TT	Màu Tóc	Chiều Cao	Cân Nặng	Dùng Thuốc?	Kết Quả
1	Đen	Tầm thước	Nhẹ	Không	Bị rám
2	Đen	Cao	Vừa phải	Có	Không
3	Râm	Thấp	Vừa phải	Có	Không
4	Đen	Thấp	Vừa phải	Không	Bị rám
5	Bạc	Tầm thước	Nặng	Không	Bị rám
6	Râm	Cao	Nặng	Không	Không
7	Râm	Tầm thước	Nặng	Không	Không
8	Đen	Thấp	Nhẹ	Có	Không

So sánh kết quả.

**Bài 4.** Sử dụng cây định danh để tìm các luật phân lớp từ bảng quyết định sau đây:

#	Trời	Áp Suất	Gió	Kết quả
1	Trong	Cao	Bắc	Không mưa
2	Mây	Cao	Nam	Mưa
3	Mây	Trung bình	Bắc	Mưa
4	Trong	Thấp	Bắc	Không mưa
5	Mây	Thấp	Bắc	Mưa
6	Mây	Cao	Bắc	Mưa
7	Mây	Thấp	Nam	Không mưa
8	Trong	Cao	Nam	Không mưa

Bạn có suy nghĩ gì về việc dùng luật kết hợp để làm luật phân lớp.

Bảng dữ liệu lúc đó sẽ có các cột <Trời, Trong>, <Trời, mây>, <Ápsuất, Cao> <Ápsuất, trung bình>, <Ápsuất, Thấp>

1. **Bài 5.** Cho bảng quyết định sau (4 điểm)

	Vóc dáng	Quốc tịch	Gia cảnh	Nhóm
--	----------	-----------	----------	------

O1	Nhỏ	Đức	Độc thân	A
O2	Lớn	Pháp	Độc thân	A
O3	Lớn	Đức	Độc thân	A
O4	Nhỏ	Ý	Độc thân	B
O5	Lớn	Đức	Có gia đình	B
O6	Lớn	Ý	Độc thân	B
O7	Lớn	Ý	Có gia đình	B
O8	Nhỏ	Đức	Có gia đình	B

- a. Tìm các luật phân lớp của bảng quyết định trên với
  - Tập thuộc tính điều kiện là { Vóc dáng, Quốc tịch, Gia cảnh }
  - Thuộc tính phân lớp là { Nhóm }
- b. Tìm các reducts bảng quyết định trên và liệt kê các luật phân lớp có số thuộc tính vế trái nhỏ nhất

**Bài 6. Sinh viên làm thực hành trên các bộ dữ liệu lớn lấy từ UCI:**

“<https://archive.ics.uci.edu/ml/datasets.html>”