

TÀI LIỆU LÝ THUYẾT TRÍ TUỆ NHÂN TẠO

Chủ đề 9 (Phần 2)

**CÁC PHƯƠNG PHÁP HỌC MÁY**

Giảng viên: ThS. Vũ Thanh Hưng

Email: [vthung@fit.hcmus.edu.vn](mailto:vthung@fit.hcmus.edu.vn)

Biên soạn: ThS. Nguyễn Ngọc Thảo

# NỘI DUNG

---

- Giới thiệu học máy
  - Định nghĩa học và học máy
  - Các vấn đề của học máy
- Các phương pháp học máy
  - Phương pháp thống kê xác suất Naïve Bayes
  - Phương pháp cây quyết định ID3
  - Phương pháp học qui nạp ILA
- Tài liệu tham khảo



---

# PHƯƠNG PHÁP CÂY QUYẾT ĐỊNH ID3

# BÀI TẬP – Xây dựng luật quyết định

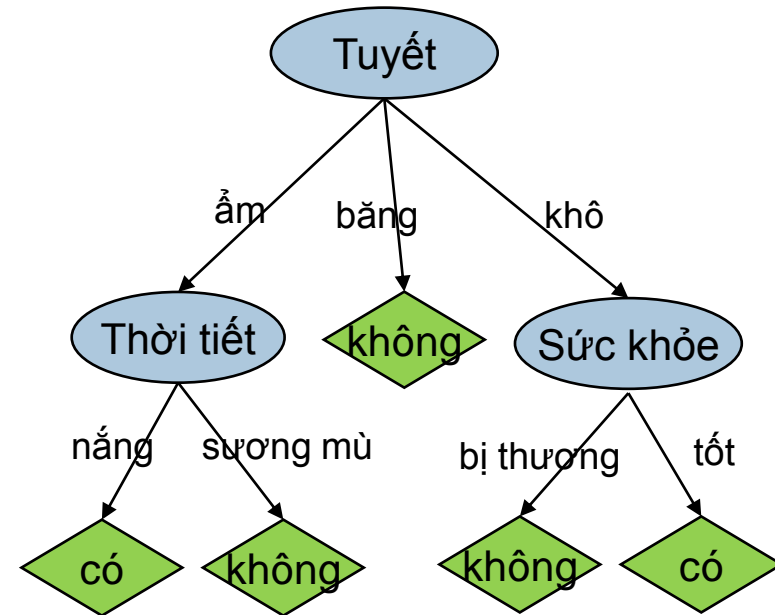
Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	ẩm	sương mù	vắng	tốt	không
2	khô	nắng	vắng	bị thương	không
3	khô	nắng	vắng	tốt	có
4	khô	nắng	cao điểm	tốt	có
5	khô	nắng	vừa phải	tốt	có
6	băng	gió	cao điểm	mệt mỏi	không
7	ẩm	nắng	vắng	tốt	có
8	băng	sương mù	vừa phải	tốt	không
9	khô	gió	vắng	tốt	có
10	khô	gió	vắng	tốt	có
11	khô	sương mù	vắng	tốt	có
12	khô	sương mù	vắng	tốt	có
13	ẩm	nắng	vừa phải	tốt	có
14	băng	sương mù	vắng	bị thương	không

Ví dụ:

IF Tuyết = băng THEN Trượt tuyết = không.

# MÔ TẢ CÂY QUYẾT ĐỊNH

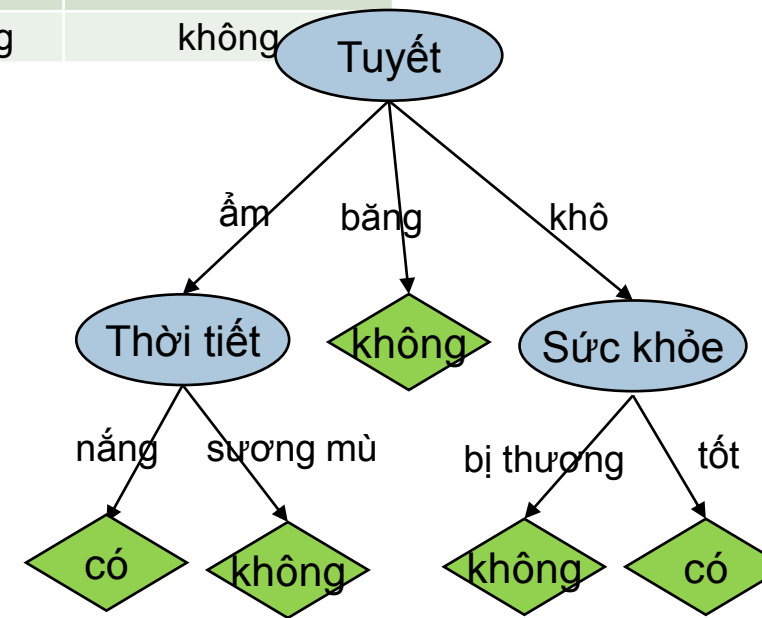
- Cấu trúc cây quyết định:
  - Mỗi nút trong kiểm tra một thuộc tính
  - Mỗi nhánh tương ứng với giá trị thuộc tính
  - Mỗi nút lá được gán một phân lớp



- Nguyên lý Occam Razor: những cây đơn giản là những cây quyết định tốt hơn

# MÔ TẢ CÂY QUYẾT ĐỊNH

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	ẩm	sương mù	vắng	tốt	không
2	khô	nắng	vắng	bị thương	không
3	khô	nắng	vắng	tốt	có
4	khô	nắng	cao điểm	tốt	có
5	khô	nắng	vừa phải	tốt	có
6	băng	gió	cao điểm	mệt mỏi	không
7	ẩm	nắng	vắng	tốt	có
8	băng	sương mù	vừa phải	tốt	không
9	khô	gió	vắng	tốt	có
10	khô	gió	vắng	tốt	có
11	khô	sương mù	vắng	tốt	có
12	khô	sương mù	vắng	tốt	có
13	ẩm	nắng	vừa phải	tốt	có
14	băng	sương mù	vắng	bị thương	không



# THUẬT TOÁN ID3

---

- Được phát triển đồng thời bởi Quinlan trong lĩnh vực Trí tuệ nhân tạo và Breiman, Friedman, Olsen và Stone trong Thống kê
- Lặp:
  1. Chọn  $A \leftarrow$  thuộc tính quyết định “tốt nhất” cho nút kế tiếp
  2. Gán  $A$  là thuộc tính quyết định cho nút
  3. Với mỗi giá trị của  $A$ , tạo nhánh con mới của nút
  4. Phân loại các mẫu huấn luyện cho các nút lá
  5. Nếu các mẫu huấn luyện được phân loại hoàn toàn thì **NGƯNG**, Ngược lại, lặp với các nút lá mới.
- Thuộc tính tốt nhất là gì?

# Chọn thuộc tính tốt nhất

---

- Kiểm tra tất cả các thuộc tính
  - Tính Average Entropy (AE) cho mỗi thuộc tính
  - Chọn thuộc tính có AE nhỏ nhất
- Cách tính AE cho mỗi thuộc tính:
  - Mỗi thuộc tính có nhiều giá trị
  - Tính Entropy (H) cho mỗi giá trị
  - AE được tính dựa vào H của các giá trị



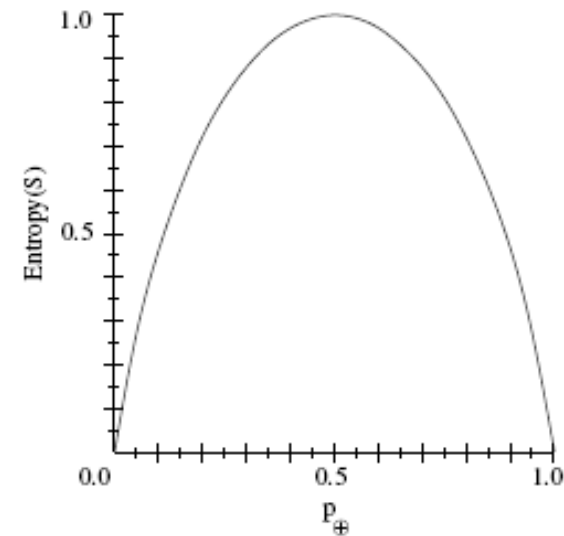
# Ví dụ: tính AE cho Tuyết

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	ẩm	sương mù	vắng	tốt	không
2	khô	nắng	vắng	bị thương	không
3	khô	nắng	vắng	tốt	có
4	khô	nắng	cao điểm	tốt	có
5	khô	nắng	vừa phải	tốt	có
6	băng	gió	cao điểm	mệt mỏi	không
7	ẩm	nắng	vắng	tốt	có
8	băng	sương mù	vừa phải	tốt	không
9	khô	gió	vắng	tốt	có
10	khô	gió	vắng	tốt	có
11	khô	sương mù	vắng	tốt	có
12	khô	sương mù	vắng	tốt	có
13	ẩm	nắng	vừa phải	tốt	có
14	băng	sương mù	vắng	bị thương	không

- Tính H cho các giá trị của Tuyết gồm {băng, ẩm, khô}
- Tính  $H_{\text{băng}}$

# ĐỘ ĐO ENTROPY

- Gọi  $S$  là tập các mẫu huấn luyện,  $p$  là tỷ lệ các mẫu dương trong  $S$ .
- $H \equiv -p \cdot \log_2 p - (1 - p) \cdot \log_2(1 - p)$



# Ví dụ: tính AE cho Tuyết

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	ẩm	sương mù	vắng	tốt	không
2	khô	nắng	vắng	bị thương	không
3	khô	nắng	vắng	tốt	có
4	khô	nắng	cao điểm	tốt	có
5	khô	nắng	vừa phải	tốt	có
6	băng	gió	cao điểm	mệt mỏi	không
7	ẩm	nắng	vắng	tốt	có
8	băng	sương mù	vừa phải	tốt	không
9	khô	gió	vắng	tốt	có
10	khô	gió	vắng	tốt	có
11	khô	sương mù	vắng	tốt	có
12	khô	sương mù	vắng	tốt	có
13	ẩm	nắng	vừa phải	tốt	có
14	băng	sương mù	vắng	bị thương	không

- S là tập các mẫu có Tuyết = băng
- $p$  là tỷ lệ các mẫu dương (Trượt tuyết = có) trong S (= 0/3)
- $H \equiv -p \cdot \log_2 p - (1 - p) \cdot \log_2 (1 - p)$
- Tính  $H_{\text{băng}}$

$$\begin{aligned}
 H_{\text{băng}} &= -0/3 \log_2(0/3) - (1-0/3) \log_2(1-0/3) \\
 &= -0/3 \log_2 0/3 - 3/3 \log_2 3/3 = 0
 \end{aligned}$$

# Ví dụ: tính AE cho Tuyết

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	ẩm	sương mù	vắng	tốt	không
2	khô	nắng	vắng	bị thương	không
3	khô	nắng	vắng	tốt	có
4	khô	nắng	cao điểm	tốt	có
5	khô	nắng	vừa phải	tốt	có
6	băng	gió	cao điểm	mệt mỏi	không
7	ẩm	nắng	vắng	tốt	có
8	băng	sương mù	vừa phải	tốt	không
9	khô	gió	vắng	tốt	có
10	khô	gió	vắng	tốt	có
11	khô	sương mù	vắng	tốt	có
12	khô	sương mù	vắng	tốt	có
13	ẩm	nắng	vừa phải	tốt	có
14	băng	sương mù	vắng	bị thương	không

- $H \equiv -p.\log_2 p - (1 - p).\log_2(1 - p)$

- Tính  $H_{\text{ẩm}}, H_{\text{khô}}$

$$p_{\text{ẩm}} = ? \Rightarrow H_{\text{ẩm}} = ?$$

$$p_{\text{khô}} = ? \Rightarrow H_{\text{khô}} = ?$$

# Chọn thuộc tính tốt nhất

---

- Kiểm tra tất cả các thuộc tính
  - Tính Average Entropy (AE) cho mỗi thuộc tính
  - Chọn thuộc tính có AE nhỏ nhất
- Cách tính AE cho mỗi thuộc tính:
  - Mỗi thuộc tính có nhiều giá trị
  - Tính Entropy (H) cho mỗi giá trị
  - **AE được tính dựa vào H của các giá trị**

# Average Entropy

---

- **Average Entropy của thuộc tính** là trung bình entropy (H) của các giá trị của thuộc tính đó.

$$AverageEntropy(A) = \sum_{v \in Value(A)} p_v H_{A=v}$$

- **Thuộc tính tốt nhất** là thuộc tính có AE nhỏ nhất → sử dụng thuộc tính này để tách cây quyết định.

# Average Entropy

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	ẩm	sương mù	vắng	tốt	không
2	khô	nắng	vắng	bị thương	không
3	khô	nắng	vắng	tốt	có
4	khô	nắng	cao điểm	tốt	có
5	khô	nắng	vừa phải	tốt	có
6	băng	gió	cao điểm	mệt mỏi	không
7	ẩm	nắng	vắng	tốt	có
8	băng	sương mù	vừa phải	tốt	không
9	khô	gió	vắng	tốt	có
10	khô	gió	vắng	tốt	có
11	khô	sương mù	vắng	tốt	có
12	khô	sương mù	vắng	tốt	có
13	ẩm	nắng	vừa phải	tốt	có
14	băng	sương mù	vắng	bị thương	không

$$Average Entropy(A) = \sum_{v \in Value(A)} p_v H_{A=v}$$

- Thuộc tính Tuyết: có 3 giá trị {băng, ẩm, khô}

→ Value(Tuyết) = {băng, ẩm, khô}

- $p_{v = \text{băng}} = 3/14$        $H_{\text{băng}} = 0$
- $p_{v = \text{ẩm}} = 3/14$        $H_{\text{ẩm}} = 0.918$
- $p_{v = \text{khô}} = 8/14$        $H_{\text{khô}} = 0.544$

# Average Entropy

$$Average Entropy(A) = \sum_{v \in Value(A)} p_v H_{A=v}$$

- Thuộc tính Tuyệt: có 3 giá trị {băng, ẩm, khô}  
→  $Value(Tuyệt) = \{băng, ẩm, khô\}$

$$p_{v = băng} = 3/14$$

$$H_{băng} = 0$$

$$p_{v = ẩm} = 3/14$$

$$H_{ẩm} = 0.918$$

$$p_{v = khô} = 8/14$$

$$H_{khô} = 0.544$$

- $AE_{Tuyệt} = 3/14 * 0 + 3/14 * 0.918 + 8/14 * 0.544$   
 $= 0.508$



# BÀI TẬP VÍ DỤ

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	ẩm	sương mù	vắng	tốt	không
2	khô	nắng	vắng	bị thương	không
3	khô	nắng	vắng	tốt	có
4	khô	nắng	cao điểm	tốt	có
5	khô	nắng	vừa phải	tốt	có
6	băng	gió	cao điểm	mệt mỏi	không
7	ẩm	nắng	vắng	tốt	có
8	băng	sương mù	vừa phải	tốt	không
9	khô	gió	vắng	tốt	có
10	khô	gió	vắng	tốt	có
11	khô	sương mù	vắng	tốt	có
12	khô	sương mù	vắng	tốt	có
13	ẩm	nắng	vừa phải	tốt	có
14	băng	sương mù	vắng	bị thương	không

# CHỌN THUỘC TÍNH LẦN 1

## Tuyết

băng

ấm

khô

0+,3-

2+,1-

7+,1-

$$H_{\text{băng}} = - 0/3 \cdot \log_2 0/3 - 3/3 \cdot \log_2 3/3 = 0$$

$$H_{\text{ấm}} = - 2/3 \cdot \log_2 2/3 - 1/3 \cdot \log_2 1/3 = 0.918$$

$$H_{\text{khô}} = - 7/8 \cdot \log_2 7/8 - 1/8 \cdot \log_2 1/8 = 0.544$$

$$\begin{aligned} AE &= 3/14 \cdot 0 + 3/14 \cdot 0.918 + 8/14 \cdot 0.544 \\ &= 0.508 \end{aligned}$$

## Thời tiết

gió

nắng

sương mù

2+,1-

5+,1-

2+,3-

$$H_{\text{gió}} = - 2/3 \cdot \log_2 2/3 - 1/3 \cdot \log_2 1/3 = 0.918$$

$$H_{\text{nắng}} = - 5/6 \cdot \log_2 5/6 - 1/6 \cdot \log_2 1/6 = 0.65$$

$$H_{\text{sm}} = - 2/5 \cdot \log_2 2/5 - 3/5 \cdot \log_2 3/5 = 0.971$$

$$\begin{aligned} AE &= 3/14 \cdot 0.918 + 6/14 \cdot 0.65 + 5/14 \cdot 0.971 \\ &= 0.822 \end{aligned}$$

# CHỌN THUỘC TÍNH LẦN 1

## Mùa

cao điểm      vắng      vừa phải

1+,1-

6+,3-

2+,1-

$$H_{\text{cđ}} = -1/2 \cdot \log_2 1/2 - 1/2 \cdot \log_2 1/2 = 1$$

$$H_{\text{vắng}} = -6/8 \cdot \log_2 6/8 - 3/8 \cdot \log_2 3/8 = 0.918$$

$$H_{\text{vp}} = -2/3 \cdot \log_2 2/3 - 1/3 \cdot \log_2 1/3 = 0.918$$

$$\begin{aligned} AE &= 2/14 \cdot 1 + 9/14 \cdot 0.918 + 3/14 \cdot 0.918 \\ &= 0.93 \end{aligned}$$

## Sức khỏe

bị thương      mệt mỏi      tốt

0+,2-

0+,1-

9+,2-

$$H_{\text{bị thương}} = -0/2 \cdot \log_2 0/2 - 2/2 \cdot \log_2 2/2 = 0$$

$$H_{\text{mệt mỏi}} = -0/1 \cdot \log_2 0/1 - 1/1 \cdot \log_2 1/1 = 0$$

$$\begin{aligned} H_{\text{tốt}} &= -9/11 \cdot \log_2 9/11 - 2/11 \cdot \log_2 2/11 \\ &= 0.684 \end{aligned}$$

$$\begin{aligned} AE &= 2/14 \cdot 0 + 1/14 \cdot 0 + 11/14 \cdot 0.684 \\ &= 0.537 \end{aligned}$$

# Chọn thuộc tính tốt nhất

---

- Kiểm tra tất cả các thuộc tính
  - Tính Average Entropy (AE) cho mỗi thuộc tính
  - **Chọn thuộc tính có AE nhỏ nhất**
- Cách tính AE cho mỗi thuộc tính:
  - Mỗi thuộc tính có nhiều giá trị
  - Tính Entropy (H) cho mỗi giá trị
  - AE được tính dựa vào H của các giá trị

# Chọn thuộc tính tốt nhất

---

- Kiểm tra tất cả các thuộc tính
  - Tính Average Entropy (AE) cho mỗi thuộc tính
  - **Chọn thuộc tính có AE nhỏ nhất**

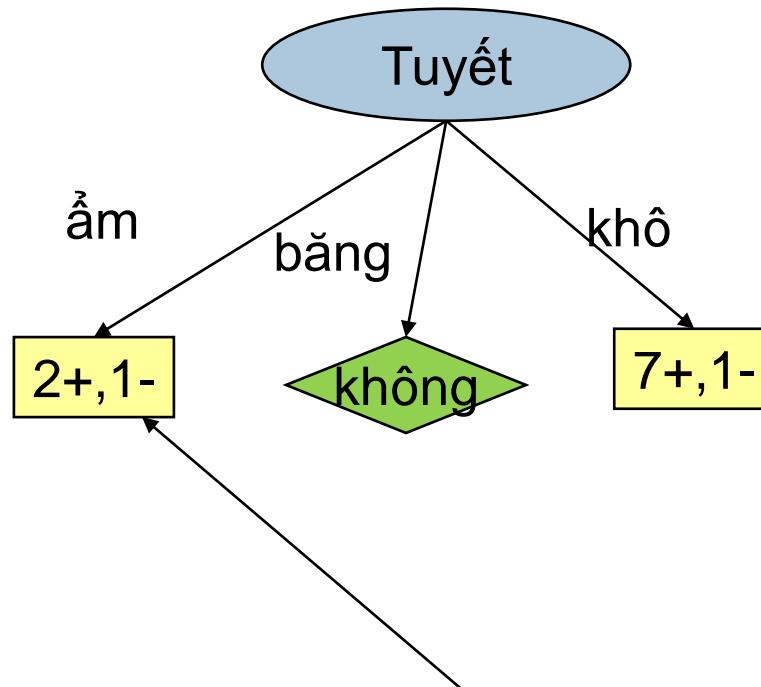
$$AE_{\text{Tuyết}} = 0.508$$

$$AE_{\text{Thời tiết}} = 0.822$$

$$AE_{\text{Mùa}} = 0.93$$

$$AE_{\text{Sức khỏe}} = 0.537$$

# CÂY QUYẾT ĐỊNH ID3



- Chọn thuộc tính gì tiếp theo?
- Tiếp tục thực hiện việc phân chia.
- Chỉ xem xét thuộc tính Thời tiết, Mùa và Sức khỏe (do Tuyết đã được dùng để phân nhánh rồi)

# Nhánh Tuyết = ẩm

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	ẩm	sương mù	vắng	tốt	không
2	khô	nắng	vắng	bị thương	không
3	khô	nắng	vắng	tốt	có
4	khô	nắng	cao điểm	tốt	có
5	khô	nắng	vừa phải	tốt	có
6	băng	gió	cao điểm	mệt mỏi	không
7	ẩm	nắng	vắng	tốt	có
8	băng	sương mù	vừa phải	tốt	không
9	khô	gió	vắng	tốt	có
10	khô	gió	vắng	tốt	có
11	khô	sương mù	vắng	tốt	có
12	khô	sương mù	vắng	tốt	có
13	ẩm	nắng	vừa phải	tốt	có
14	băng	sương mù	vắng	bị thương	không



Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	ẩm	sương mù	vắng	tốt	không
7	ẩm	nắng	vắng	tốt	có
13	ẩm	nắng	vừa phải	tốt	có

# Nhánh Tuyết = ẩm

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	ẩm	sương mù	vắng	tốt	không
7	ẩm	nắng	vắng	tốt	có
13	ẩm	nắng	vừa phải	tốt	có

**Thời tiết**

nắng

sương mù

2+,0-

0+,1-

**Mùa**

vắng

vừa phải

1+,1-

1+,0-

**Sức khỏe**

tốt

2+,1-

$$H_{\text{nắng}} = -2/2 \log_2 2/2 - 0/2 \log_2 0/2 = 0$$

$$H_{\text{sm}} = -0/1 \log_2 0/1 - 1/1 \log_2 1/1 = 0$$

$$AE = 2/3 H_{\text{nắng}} + 1/3 H_{\text{sương mù}} = 0$$

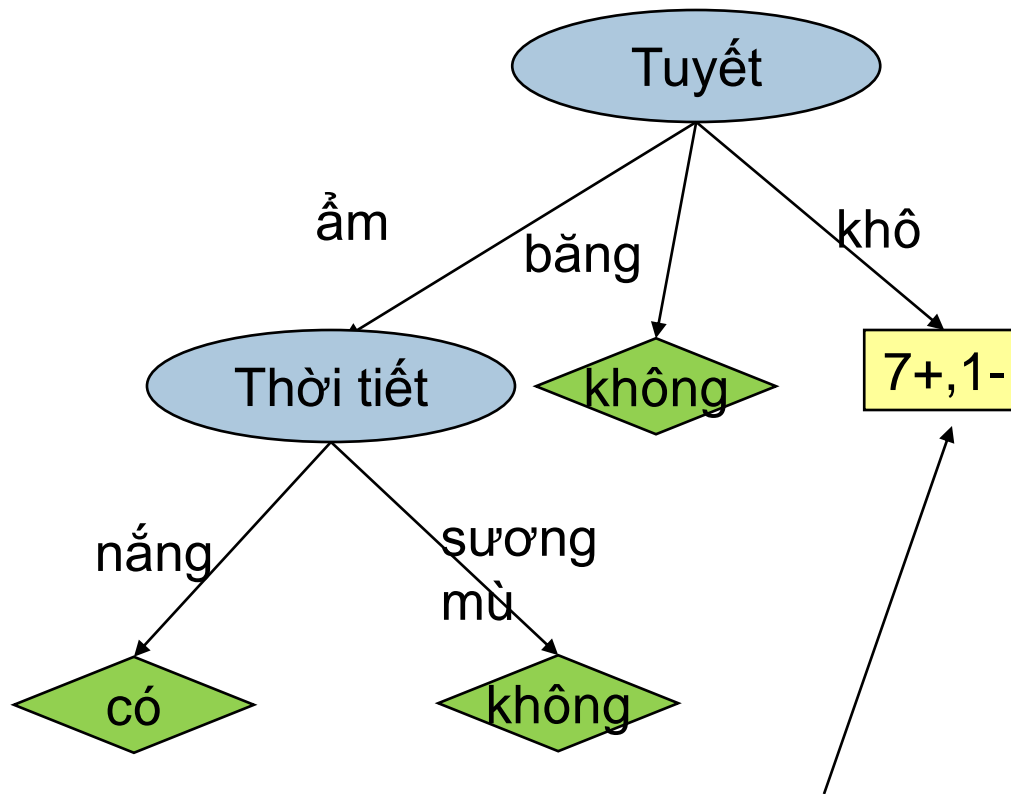


# Nhánh Tuyết = ẩm

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	ẩm	sương mù	vắng	tốt	không
7	ẩm	nắng	vắng	tốt	có
13	ẩm	nắng	vừa phải	tốt	có

Thời tiết		Mùa		Sức khỏe
nắng	sương mù	vắng	vừa phải	tốt
2+,0-	0+,1-	1+,1-	1+,0-	2+,1-
$H_{\text{nắng}} = 0$		$H_{\text{vắng}} = 1$		$H_{\text{tốt}} = 0.918$
$H_{\text{sm}} = 0$		$H_{\text{vp}} = 0$		$AE = 3/3 \cdot 0.918$
$AE = 0$		$AE = 2/3 \cdot 1 + 1/3 \cdot 0 = 0.667$		$= 0.918$

# CÂY QUYẾT ĐỊNH ID3



- Chọn thuộc tính gì tiếp theo?
- Tiếp tục thực hiện việc phân chia.
- Chỉ xem xét thuộc tính Thời tiết, Mùa và Sức khỏe (do Tuyết đã được dùng để phân nhánh rồi)

# Nhánh Tuyết = khô

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	ẩm	sương mù	vắng	tốt	không
2	khô	nắng	vắng	bị thương	không
3	khô	nắng	vắng	tốt	có
4	khô	nắng	cao điểm	tốt	có
5	khô	nắng	vừa phải	tốt	có
6	băng	gió	cao điểm	mệt mỏi	không
7	ẩm	nắng	vắng	tốt	có
8	băng	sương mù	vừa phải	tốt	không
9	khô	gió	vắng	tốt	có
10	khô	gió	vắng	tốt	có
11	khô	sương mù	vắng	tốt	có
12	khô	sương mù	vắng	tốt	có
13	ẩm	nắng	vừa phải	tốt	có
14	băng	sương mù	vắng	bị thương	không

# Nhánh Tuyết = khô

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
2	khô	nắng	vắng	bị thương	không
3	khô	nắng	vắng	tốt	có
4	khô	nắng	cao điểm	tốt	có
5	khô	nắng	vừa phải	tốt	có
9	khô	gió	vắng	tốt	có
10	khô	gió	vắng	tốt	có
11	khô	sương mù	vắng	tốt	có
12	khô	sương mù	vắng	tốt	có

# Nhánh Tuyết = khô

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
2	khô	nắng	vắng	bị thương	không
3	khô	nắng	vắng	tốt	có
4	khô	nắng	cao điểm	tốt	có
5	khô	nắng	vừa phải	tốt	có
9	khô	gió	vắng	tốt	có
10	khô	gió	vắng	tốt	có
11	khô	sương mù	vắng	tốt	có
12	khô	sương mù	vắng	tốt	có

## Thời tiết

gió	nắng	sương mù
?+,?-	?+,?-	?+,?-

$$H_{\text{gió}} = ?$$

$$H_{\text{nắng}} = ?$$

$$H_{\text{sm}} = ?$$

$$AE = ?$$

## Mùa

cao điểm	vắng	vừa phải
?+,?-	?+,?-	?+,?-

$$H_{\text{cđ}} = ?$$

$$H_{\text{vắng}} = ?$$

$$H_{\text{vp}} = ?$$

$$AE = ?$$

## Sức khỏe

bị thương	tốt
?+,?-	?+,?-

$$H_{\text{tốt}} = ?$$

$$H_{\text{bị thương}} = ?$$

$$AE = ?$$

# Nhánh Tuyết = khô

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
2	khô	nắng	vắng	bị thương	không
3	khô	nắng	vắng	tốt	có
4	khô	nắng	cao điểm	tốt	có
5	khô	nắng	vừa phải	tốt	có
9	khô	gió	vắng	tốt	có
10	khô	gió	vắng	tốt	có
11	khô	sương mù	vắng	tốt	có
12	khô	sương mù	vắng	tốt	có

## Thời tiết

gió	nắng	sương mù
2+,0-	3+,1-	2+,0-

$$H_{\text{gió}} = 0$$

$$H_{\text{nắng}} = 0.811$$

$$H_{\text{sm}} = 0$$

$$AE = 4/8 * 0.811 = 0.401$$

## Mùa

cao điểm	vắng	vừa phải
1+,0-	5+,1-	1+,0-

$$H_{\text{cđ}} = 0$$

$$H_{\text{vắng}} = 0.650$$

$$H_{\text{vp}} = 0$$

$$AE = 6/8 * 0.650 = 0.488$$

## Sức khỏe

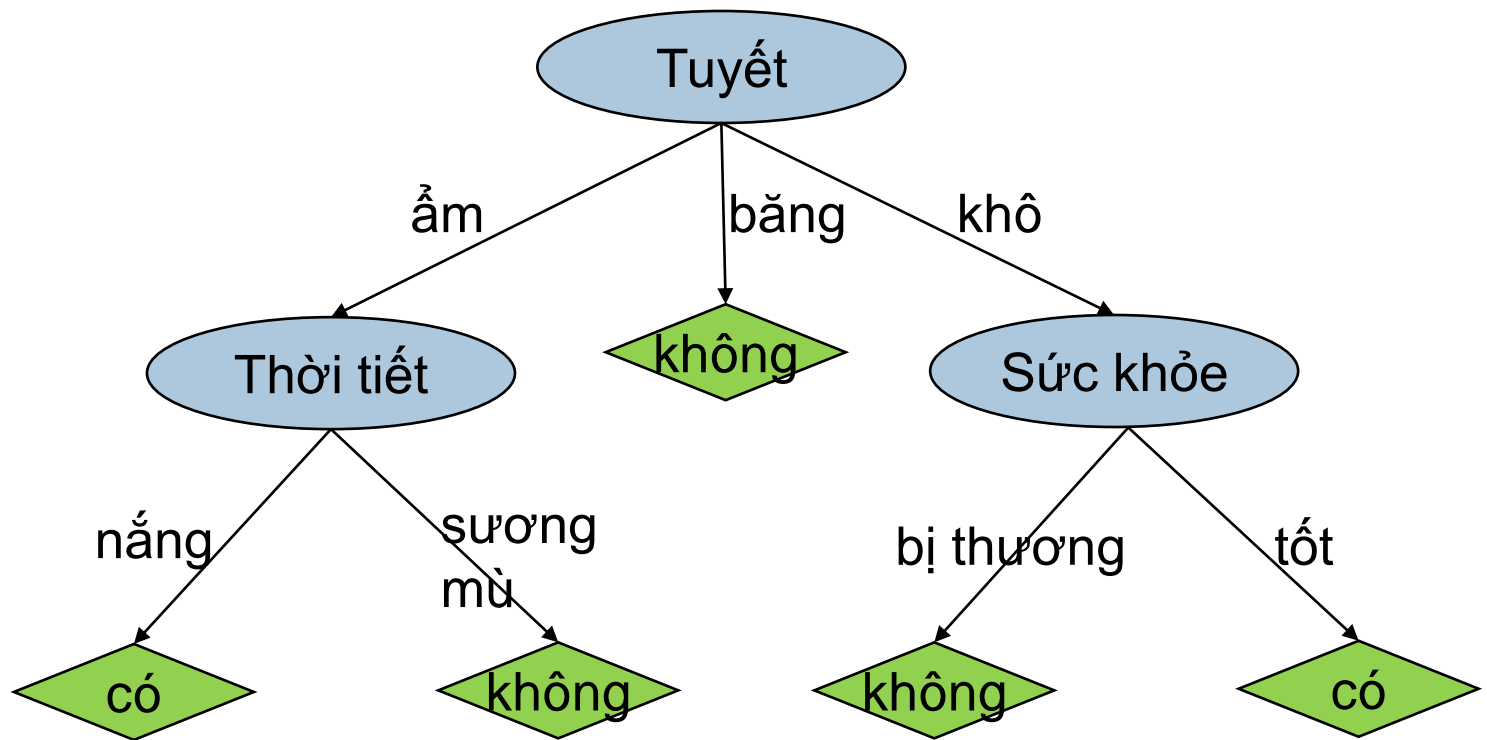
bị thương	tốt
0+,1-	7+,0-

$$H_{\text{tốt}} = 0$$

$$H_{\text{bị thương}} = 0$$

$$AE = 0$$

# CÂY QUYẾT ĐỊNH ID3



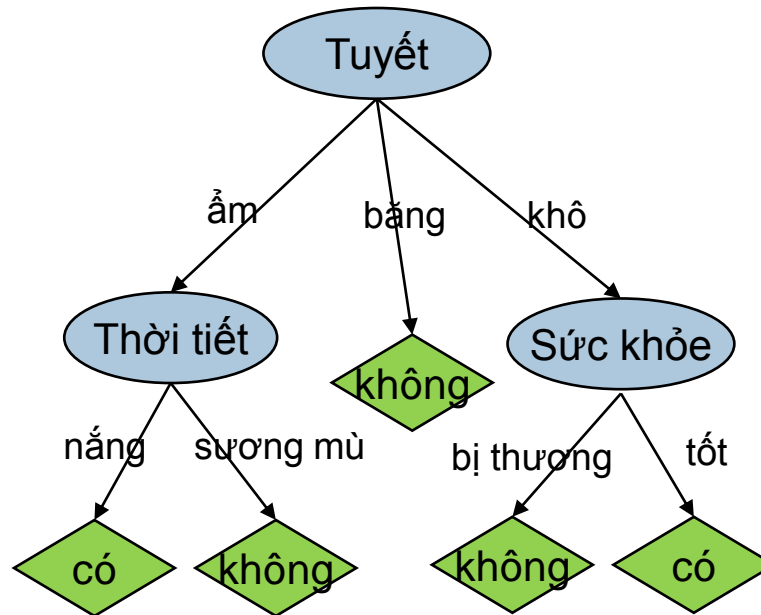
# TRI THỨC DẠNG LUẬT

---

- Tri thức được biểu diễn dưới dạng luật:  
IF Điều kiện 1  $\wedge$  Điều kiện 2...  
THEN Kết luận
- Dễ hiểu với con người, được sử dụng chủ yếu trong các hệ chuyên gia
- Rút luật từ cây quyết định: đi từ nút gốc đến nút lá, lấy các phép kiểm tra làm tiền đề và phân loại của nút lá làm kết quả



# TRI THỨC DẠNG LUẬT



- IF Tuyết = ẩm AND Thời tiết = nắng THEN Trượt tuyết = có.
- IF Tuyết = ẩm AND Thời tiết = sương mù THEN Trượt tuyết = có.
- IF Tuyết = băng THEN Trượt tuyết = không.
- IF Tuyết = khô AND Sức khỏe = bị thương THEN Trượt tuyết = không.
- IF Tuyết = khô AND Sức khỏe = tốt THEN Trượt tuyết = có.



---

# PHƯƠNG PHÁP HỌC QUY NẠP ILA

# DỮ LIỆU MINH HỌA

Mẫu	Kích cỡ	Màu sắc	Hình dáng	Quyết định
1	Vừa	Xanh dương	Hộp	Mua
2	Nhỏ	Đỏ	Nón	Không mua
3	Nhỏ	Đỏ	Cầu	Mua
4	Lớn	Đỏ	Nón	Không mua
5	Lớn	Xanh lá	Trụ	Mua
6	Lớn	Đỏ	Trụ	Không mua
7	Lớn	Xanh lá	Cầu	Mua

# THUẬT TOÁN ILA

---

1. Chia tập mẫu thành các tập con ứng với thuộc tính quyết định.

# VÍ DỤ MINH HỌA ILA

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
1	Vừa	Xanh dương	Hộp	Mua
3	Nhỏ	Đỏ	Cầu	Mua
5	Lớn	Xanh lá	Trụ	Mua
7	Lớn	Xanh lá	Cầu	Mua

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
2	Nhỏ	Đỏ	Nón	Không mua
4	Lớn	Đỏ	Nón	Không mua
6	Lớn	Đỏ	Trụ	Không mua

# THUẬT TOÁN ILA

---

1. Chia tập mẫu thành các tập con ứng với thuộc tính quyết định.
2. Với mỗi bảng con
  - a. Với mỗi tổ hợp  $m$  thuộc tính có thể (khởi tạo  $m = 1$ )
    - i. Tìm các giá trị chỉ xuất hiện ở bảng con này mà không xuất hiện ở các bảng con khác.
    - ii. (Nếu có nhiều tổ hợp thì chọn tổ hợp có số lượng mẫu tin nhiều nhất)

# VÍ DỤ MINH HỌA ILA

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
1	Vừa	Xanh dương	Hộp	Mua
3	Nhỏ	Đỏ	Cầu	Mua
5	Lớn	Xanh lá	Trụ	Mua
7	Lớn	Xanh lá	Cầu	Mua

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
2	Nhỏ	Đỏ	Nón	Không mua
4	Lớn	Đỏ	Nón	Không mua
6	Lớn	Đỏ	Trụ	Không mua

# THUẬT TOÁN ILA

---

1. Chia tập mẫu thành các tập con ứng với thuộc tính quyết định.
2. Với mỗi bảng con
  - a. Với mỗi tổ hợp  $m$  thuộc tính có thể (khởi tạo  $m = 1$ )
    - i. Tìm các giá trị chỉ xuất hiện ở bảng con này mà không xuất hiện ở các bảng con khác.
    - ii. (Nếu có nhiều tổ hợp thì chọn tổ hợp có số lượng mẫu tin nhiều nhất)
    - iii. Sử dụng tổ hợp thuộc tính - giá trị vừa tìm để tạo luật
    - iv. Đánh dấu các dòng đã xét



# VÍ DỤ MINH HỌA ILA

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
1	Vừa	Xanh dương	Hộp	Mua
3	Nhỏ	Đỏ	Cầu	Mua
5	Lớn	Xanh lá	Trụ	Mua
7	Lớn	Xanh lá	Cầu	Mua

IF Màu sắc = Xanh lá THEN Quyết định = Mua

# THUẬT TOÁN ILA

---

1. Chia tập mẫu thành các tập con ứng với thuộc tính quyết định.
2. Với mỗi bảng con
  - a. Với mỗi tổ hợp  $m$  thuộc tính có thể (khởi tạo  $m = 1$ )
    - i. Tìm các giá trị chỉ xuất hiện ở bảng con này mà không xuất hiện ở các bảng con khác.
    - ii. (Nếu có nhiều tổ hợp thì chọn tổ hợp có số lượng mẫu tin nhiều nhất)
    - iii. Sử dụng tổ hợp thuộc tính - giá trị vừa tìm để tạo luật
    - iv. Đánh dấu các dòng đã xét
  - b. Nếu còn dòng chưa xét, lặp lại bước a

# VÍ DỤ MINH HỌA ILA

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
1	Vừa	Xanh dương	Hộp	Mua
3	Nhỏ	Đỏ	Cầu	Mua
5	Lớn	Xanh lá	Trụ	Mua
7	Lớn	Xanh lá	Cầu	Mua

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
2	Nhỏ	Đỏ	Nón	Không mua
4	Lớn	Đỏ	Nón	Không mua
6	Lớn	Đỏ	Trụ	Không mua

IF Màu sắc = Xanh lá THEN Quyết định = Mua

# VÍ DỤ MINH HỌA ILA

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
1	Vừa	Xanh dương	Hộp	Mua
3	Nhỏ	Đỏ	Cầu	Mua
5	Lớn	Xanh lá	Trụ	Mua
7	Lớn	Xanh lá	Cầu	Mua

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
2	Nhỏ	Đỏ	Nón	Không mua
4	Lớn	Đỏ	Nón	Không mua
6	Lớn	Đỏ	Trụ	Không mua

IF Màu sắc = Xanh lá THEN Quyết định = Mua

IF Kích cỡ = Vừa THEN Quyết định = Mua

# VÍ DỤ MINH HỌA ILA

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
1	Vừa	Xanh dương	Hộp	Mua
3	Nhỏ	Đỏ	Cầu	Mua
5	Lớn	Xanh lá	Trụ	Mua
7	Lớn	Xanh lá	Cầu	Mua

IF Màu sắc = Xanh lá THEN Quyết định = Mua

IF Kích cỡ = Vừa THEN Quyết định = Mua

IF Hình dáng = Cầu THEN Quyết định = Mua

# THUẬT TOÁN ILA

---

1. Chia tập mẫu thành các tập con ứng với thuộc tính quyết định.
2. Với mỗi bảng con
  - a. Với mỗi tổ hợp  $m$  thuộc tính có thể (khởi tạo  $m = 1$ )
    - i. Tìm các giá trị chỉ xuất hiện ở bảng con này mà không xuất hiện ở các bảng con khác.
    - ii. (Nếu có nhiều tổ hợp thì chọn tổ hợp có số lượng mẫu tin nhiều nhất)
    - iii. Sử dụng tổ hợp thuộc tính - giá trị vừa tìm để tạo luật
    - iv. Đánh dấu các dòng đã xét
  - b. Nếu còn dòng chưa xét, lặp lại bước a
3. Lặp lại bước 2 với các bảng con khác

# VÍ DỤ MINH HỌA ILA

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
1	Vừa	Xanh dương	Hộp	Mua
3	Nhỏ	Đỏ	Cầu	Mua
5	Lớn	Xanh lá	Trụ	Mua
7	Lớn	Xanh lá	Cầu	Mua

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
2	Nhỏ	Đỏ	Nón	Không mua
4	Lớn	Đỏ	Nón	Không mua
6	Lớn	Đỏ	Trụ	Không mua

# VÍ DỤ MINH HỌA ILA

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
1	Vừa	Xanh dương	Hộp	Mua
3	Nhỏ	Đỏ	Cầu	Mua
5	Lớn	Xanh lá	Trụ	Mua
7	Lớn	Xanh lá	Cầu	Mua

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
2	Nhỏ	Đỏ	Nón	Không mua
4	Lớn	Đỏ	Nón	Không mua
6	Lớn	Đỏ	Trụ	Không mua

IF Hình dáng = Nón THEN Quyết định = Không mua



# VÍ DỤ MINH HỌA ILA

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
1	Vừa	Xanh dương	Hộp	Mua
3	Nhỏ	Đỏ	Cầu	Mua
5	Lớn	Xanh lá	Trụ	Mua
7	Lớn	Xanh lá	Cầu	Mua

STT	Kích cỡ	Màu sắc	Hình dáng	Quyết định
2	Nhỏ	Đỏ	Nón	Không mua
4	Lớn	Đỏ	Nón	Không mua
6	Lớn	Đỏ	Trụ	Không mua

IF Hình dáng = Nón THEN Quyết định = Không mua

IF Kích cỡ = Lớn AND Màu sắc = Đỏ THEN Quyết định = Không mua

# TỔNG KẾT

---

- Hiểu khái niệm học và học máy.
- Nắm vững ý tưởng các mô hình học máy: thống kê Naïve Bayes, cây quyết định ID3 và học quy nạp ILA.

# TÀI LIỆU THAM KHẢO

---

- Tài liệu bài giảng môn học
- Chapter 6. J. Han and M. Kamber, *Data Mining: Concepts and Techniques. Second Edition*. 2006
- Nguyễn Hoàng Tú Anh. *Slide bài giảng môn Khai thác dữ liệu và Ứng dụng*. Khoa Công nghệ Thông tin, 2011.
- <http://bengio.abracadoudou.com/lectures/intro.pdf>



---

# KẾT THÚC CHỦ ĐỀ