

TÀI LIỆU LÝ THUYẾT TRÍ TUỆ NHÂN TẠO

Chủ đề 9 (Phần 1)

CÁC PHƯƠNG PHÁP HỌC MÁY

Giảng viên: ThS. Vũ Thanh Hưng

Email: 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



GIỚI THIỆU HỌC MÁY

HỌC LÀ GÌ?

- **Học** là quá trình tiếp nhận tri thức mới hoặc cập nhật tri thức về hành vi, kỹ năng,...và liên quan nhiều thông tin khác nhau.
 - Thay đổi để tốt hơn (theo một điều kiện định trước) khi có tình huống tương tự xảy ra.
 - Học không phải là “học thuộc lòng”.
- Học là một trong những khả năng quan trọng của con người.

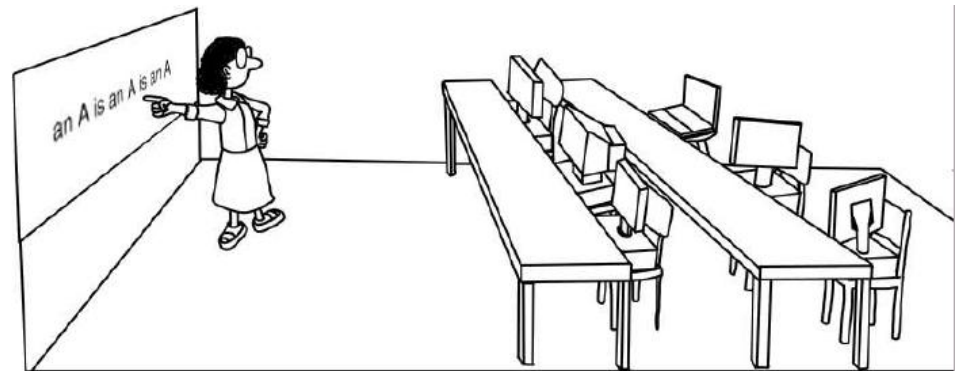
HỌC MÁY LÀ GÌ?

1. Định nghĩ bởi Arthur Samuel (1959):

Lĩnh vực nghiên cứu giúp máy tính có khả năng học không cần phải lập trình một cách tường minh.

2. Định nghĩa bởi Tom M. Mitchell :

Một chương trình máy tính được xem là “có khả năng học từ kinh nghiệm E ” ứng với nhiệm vụ T và độ đo P , nếu nó thực hiện nhiệm vụ T đó, được đo bởi P , tốt hơn kinh nghiệm E .



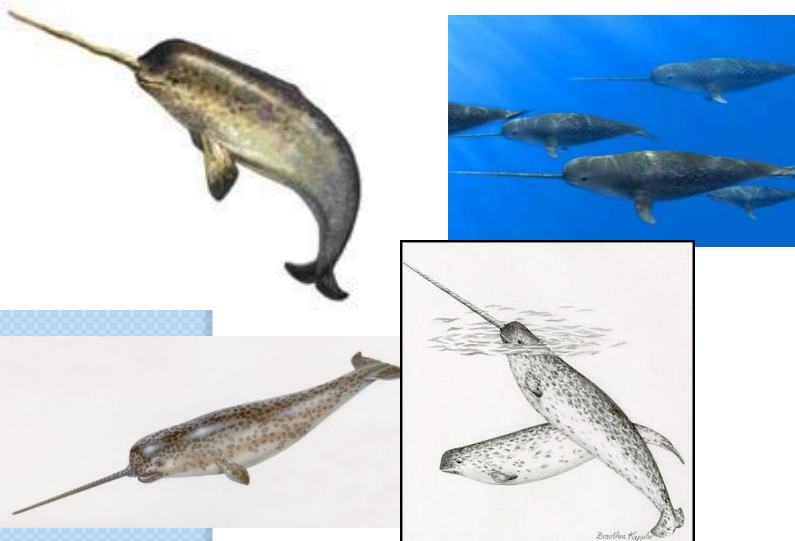
Ví dụ:

- Nhiệm vụ T: Cho bức ảnh của một loài động vật A, xác định A có phải là loài kì lân biển hay không?
- Để giải quyết được nhiệm vụ này: một người bình thường sẽ yêu cầu gì đầu tiên?

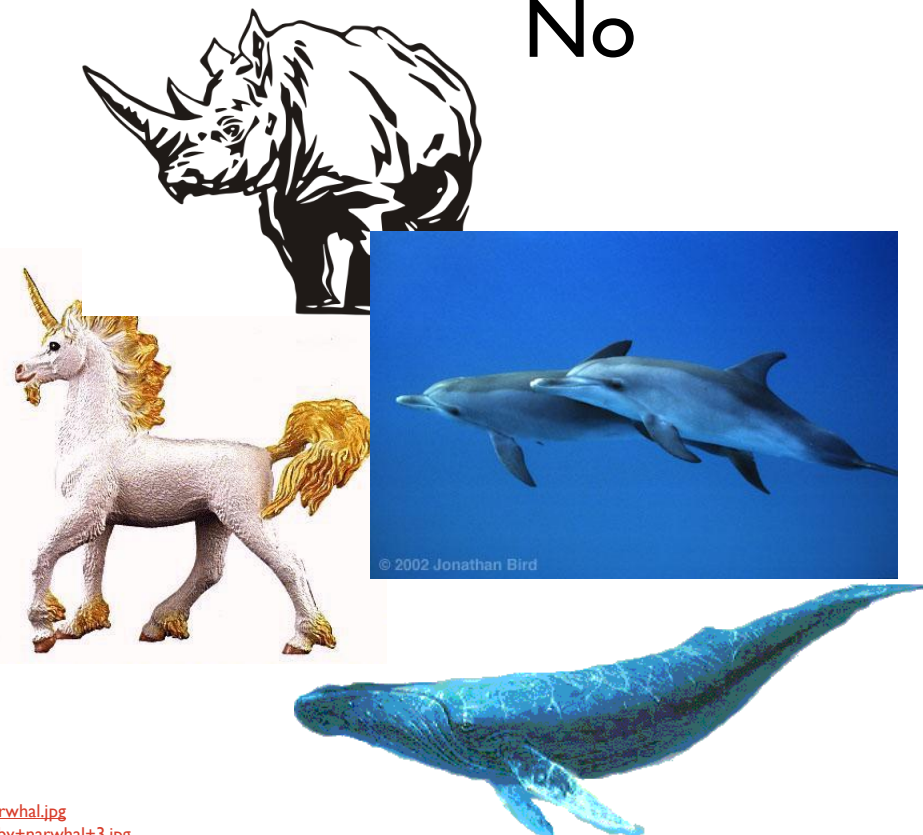
Ví dụ:

- Nhiệm vụ T: Cho bức ảnh của một loài động vật A, xác định A có phải là loài kì lân biển hay không?

Yes



No



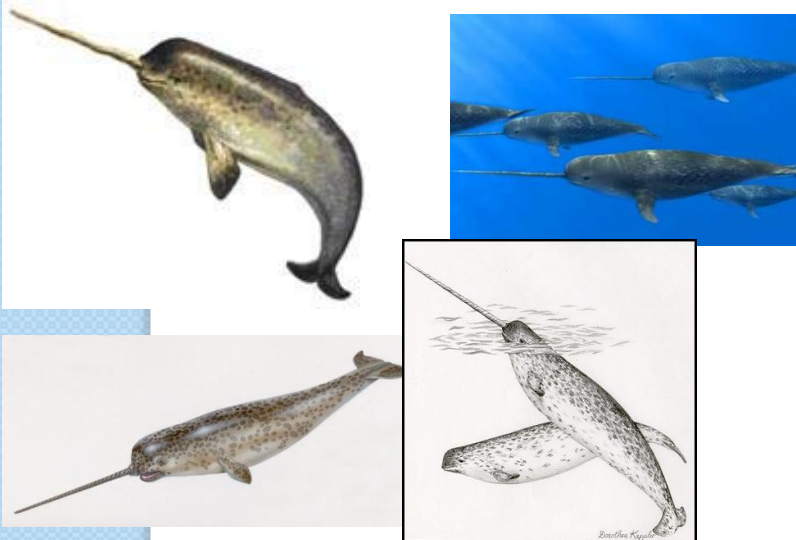
Nguồn:

<http://library.thinkquest.org/3500/images/narwhal?.jpg>
http://?bp.blogspot.com/_hYkL4prVW?M/TGsF4o?iiL/AAAAAAAAA4/aooHDiQv?Uw/s??00/narwhal.jpg
http://?bp.blogspot.com/~MSzrRSaaYXY/T8?r_boL5zl/AAAAAAAAAlas/wd57KP8LrTM/s??00/baby+narwhal+3.jpg
<http://www.pc.gc.ca/progs/amnc-nmca/systemplan/especies-species/images/c?c?0nar.jpg>
http://www.mybelvedere.com/Images/autodecals/tribalanimallarge/rino%?QTAR_3?.png
<http://www.massbayguides.com/Images/Moving%?0VWhale.gif>
<http://thegalleryofmonstertoys.com0swing/unicorn.jpg>
http://www.oceanicresearch.org/ipegs/spotted_dolphin3.jpg

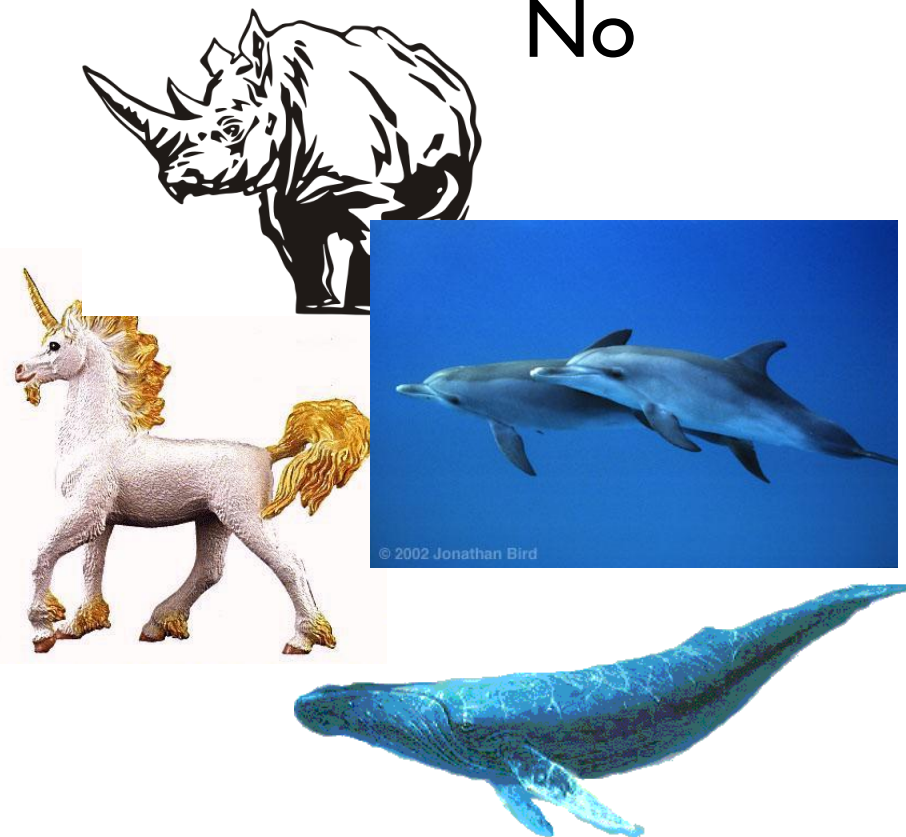
Ví dụ:

- Nhiệm vụ T: Cho bức ảnh của một loài động vật A, xác định A có phải là loài kì lân biển hay không?

Yes



No



Kinh nghiệm E

Tốt hơn kinh nghiệm E

- “Tốt hơn” cho những dữ liệu mới (không giống hoàn toàn dữ liệu cũ) vẫn có thể giải quyết tốt công việc T

Dùng kinh nghiệm E



Có trong dữ liệu



Kì lân biển



Không trong dữ liệu



??????



Không trong dữ liệu



?????

Nguồn:

<http://library.thinkquest.org/3500/images/narwhal?.jpg>

http://www.fishworldfyi.com/gallery/cache/All%?0Fish/Caroon/Tots/Whale%?0Tots/Cuddly%?0Narwhal_300.jpg

<http://graytaxidermy.com/Gallery/fishmounts/fishmount-images/swordfish-large.jpg>

Tốt hơn kinh nghiệm E

- “Tốt hơn” cho những dữ liệu mới (không giống hoàn toàn dữ liệu cũ) vẫn có thể giải quyết tốt công việc T

Tốt hơn kinh nghiệm E



Có trong dữ liệu



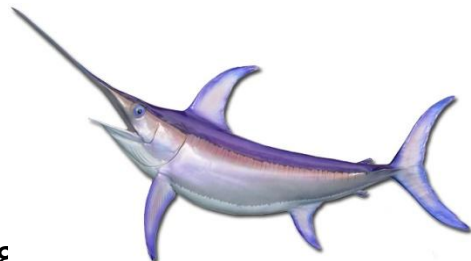
Kì lân biển



Không trong dữ liệu



Kì lân biển



Không trong dữ liệu



Không phải

Nguồn...

<http://library.thinkquest.org/3500/images/narwhal?.jpg>

http://www.fishworldfyi.com/gallery/cache/All%?0Fish/Caroon/Tots/Whale%?0Tots/Cuddly%?0Narwhal_300.jpg

<http://graytaxidermy.com/Gallery/fishmounts/fishmount-images/swordfish-large.jpg>

ỨNG DỤNG CỦA HỌC MÁY

- Ba lĩnh vực thích hợp cho học máy:
 - Khai thác dữ liệu: sử dụng dữ liệu thu thập trong quá khứ để cải thiện quyết định.
 - Ví dụ: hồ sơ bệnh án → tri thức y học, chẩn đoán mới
 - Các phần mềm ứng dụng không thể lập trình bằng thủ công.
 - Ví dụ: tự động điều khiển xe, nhận dạng giọng nói
 - Các chương trình tự điều chỉnh.
 - Ví dụ: chương trình Newsreader “học” sở thích của người dùng

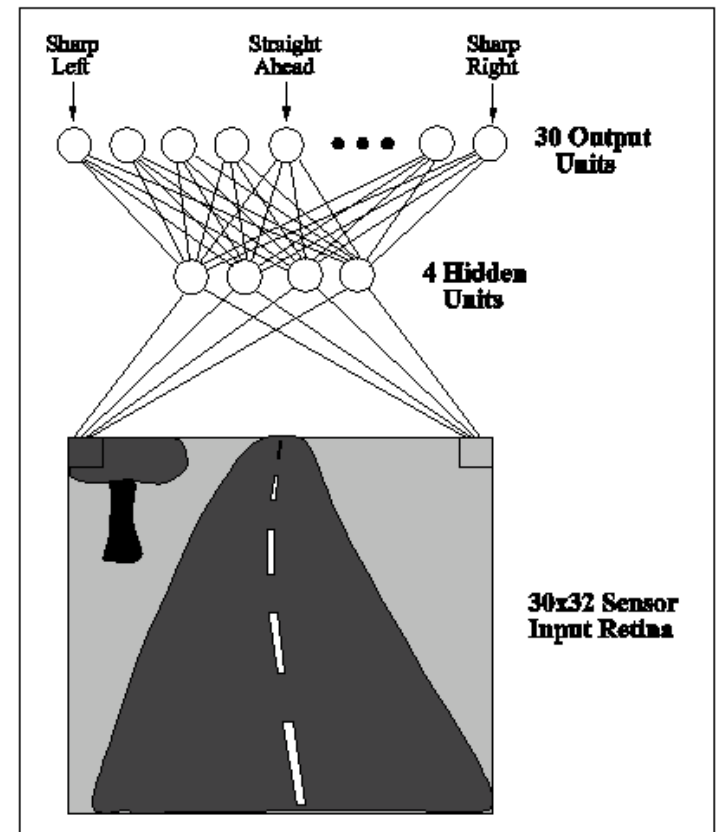
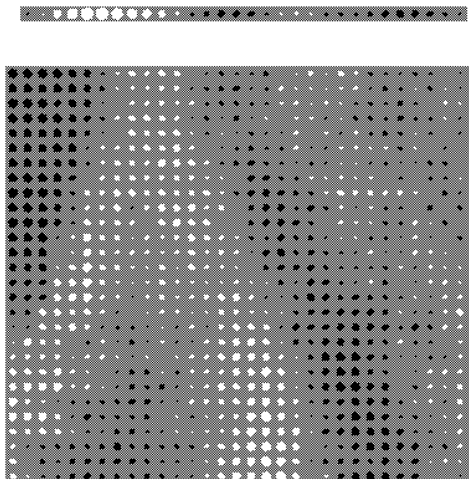
CHẨN ĐOÁN Y KHOA

- Hệ thống học kinh nghiệm từ những bệnh án đã có để dự đoán bệnh cho các ca mới.

Mẫu	Sốt	Nôn	Tiêu chảy	Rét run	Phân loại
d ₇	không	không	không	không	khỏe mạnh
d ₇	trung bình	không	không	không	cúm
d ₃	cao	không	không	có	cúm
d ₄	cao	có	có	không	ngộ độc salmonella
d ₅	trung bình	không	có	không	ngộ độc salmonella
d ₇	không	có	có	không	viêm ruột
d ₇	trung bình	có	có	không	viêm ruột
d ₈	không	không	có	có	?

HỆ THỐNG ALVINN

- Hệ thống lái xe tự động **ALVINN** (NavLab, CMU, 1989).



<http://www.ri.cmu.edu/>

ỨNG DỤNG CỦA HỌC MÁY

- Xử lý thị giác
 - Phát hiện/xác minh khuôn mặt
 - Nhận dạng chữ viết tay
- Xử lý tiếng nói
 - Nhận dạng âm vị/từ/câu/đối tượng người

ỨNG DỤNG CỦA HỌC MÁY

- Các ứng dụng khác
 - Đánh chỉ mục: Google, khai thác văn bản, truy vấn thông tin
 - Tài chính: dự đoán khả năng vay nợ, quản lý rủi ro và đầu tư
 - Truyền thông: dự đoán tình hình giao thông
 - Trò chơi: cờ vua, cờ vây
 - Điều khiển: huấn luyện robot
- ...và còn rất nhiều ứng dụng.

ỨNG DỤNG CỦA HỌC MÁY

- Học máy được ứng dụng trong trường hợp
 - Con người không thể tham gia vào quá trình.
 - Ví dụ: thám hiểm sao Hỏa
 - Con người không thể giải thích về chuyên môn.
 - Ví dụ: nhận dạng vân tay, giọng nói
 - Các giải pháp thay đổi theo thời gian.
 - Ví dụ: định tuyến trên một mạng máy tính
 - Giải pháp cần phải thích nghi với những trường hợp cụ thể.
 - Ví dụ: nhận dạng sinh trắc trong các ứng dụng bảo mật

Ứng dụng

- Dịch tự động

Google

Try a new browser with automatic translation. [Download Google Chrome](#) [Dismiss](#)

Translate

From: English - detected To: Vietnamese Translate

English Spanish French English - detected

Today, we will learn machine learning

English Spanish Vietnamese

Hôm nay, chúng ta sẽ học học máy

New! Click the words above to edit and view alternate translations. [Dismiss](#)

- Đưa lời khuyên khi mua sắm

Customers Who Bought This Item Also Bought



[Pattern Recognition and Machine Learning \(Information Science and Statistics\)](#) by Christopher M. Bishop

★★★★☆ (30) \$60.50



[Artificial Intelligence: A Modern Approach \(2nd Edition\) \(Prentice Hall Series in Artificial Intelligence\)](#) by Stuart Russell

★★★★☆ (76) \$115.00



[The Elements of Statistical Learning](#) by T. Hastie

★★★★☆ (25) \$72.20



[Pattern Classification \(2nd Edition\)](#) by Richard O. Duda

★★★★☆ (25) \$115.00



[Data Mining: Practical Machine Learning Tools and Techniques, Second Edition \(Morgan Kaufmann Series in Data Management Systems\)](#) by Ian H. Witten

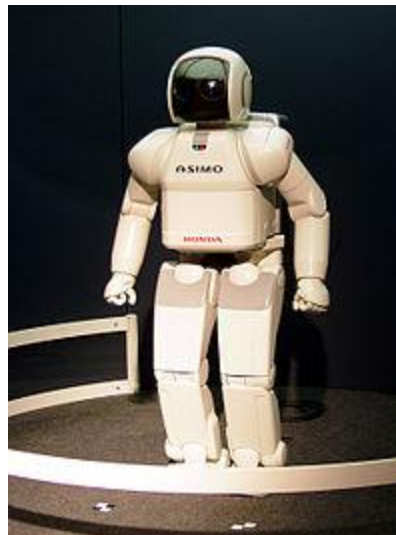
★★★★☆ (21) \$39.66

Ứng dụng

- Nhận dạng mặt người



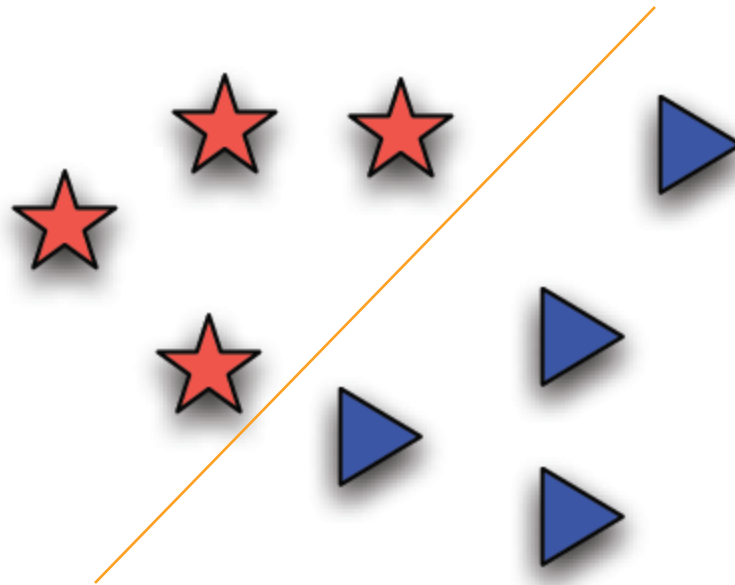
- Robot



TẠI SAO HỌC LẠI KHÓ?

- Cho trước một lượng **hữu hạn** dữ liệu huấn luyện. Ta cần suy ra một **quan hệ** trên không gian **vô hạn**.
- Trong thực tế, có vô số **quan hệ** như thế → Nên chọn cái nào?
- Ví dụ:

Kì lân biển



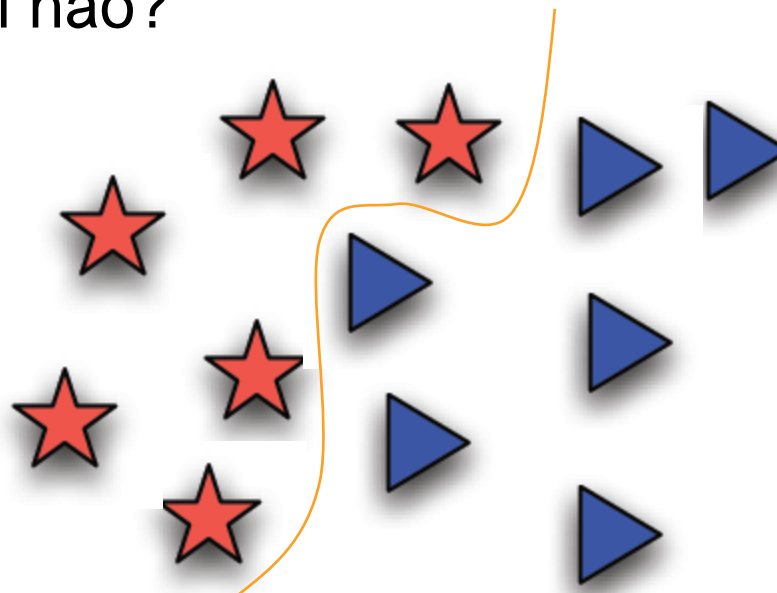
Không phải kì lân biển

Có phản ánh đúng thực tế?

TẠI SAO HỌC LẠI KHÓ?

- Cho trước một lượng **hữu hạn** dữ liệu huấn luyện. Ta cần suy ra một **quan hệ** trên không gian **vô hạn**.
- Trong thực tế, có vô số **quan hệ** như thế → Nên chọn cái nào?
- Ví dụ:

Kì lân biển



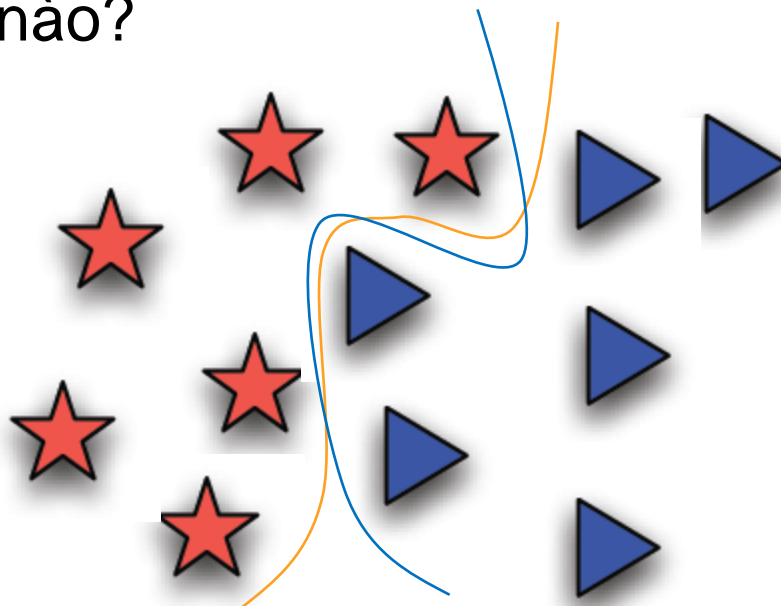
Không phải kì lân biển

- Thực tế

TẠI SAO HỌC LẠI KHÓ?

- Cho trước một lượng **hữu hạn** dữ liệu huấn luyện. Ta cần suy ra một **quan hệ** trên không gian **vô hạn**.
- Trong thực tế, có vô số **quan hệ** như thế → Nên chọn cái nào?
- Ví dụ:

Kì lân biển



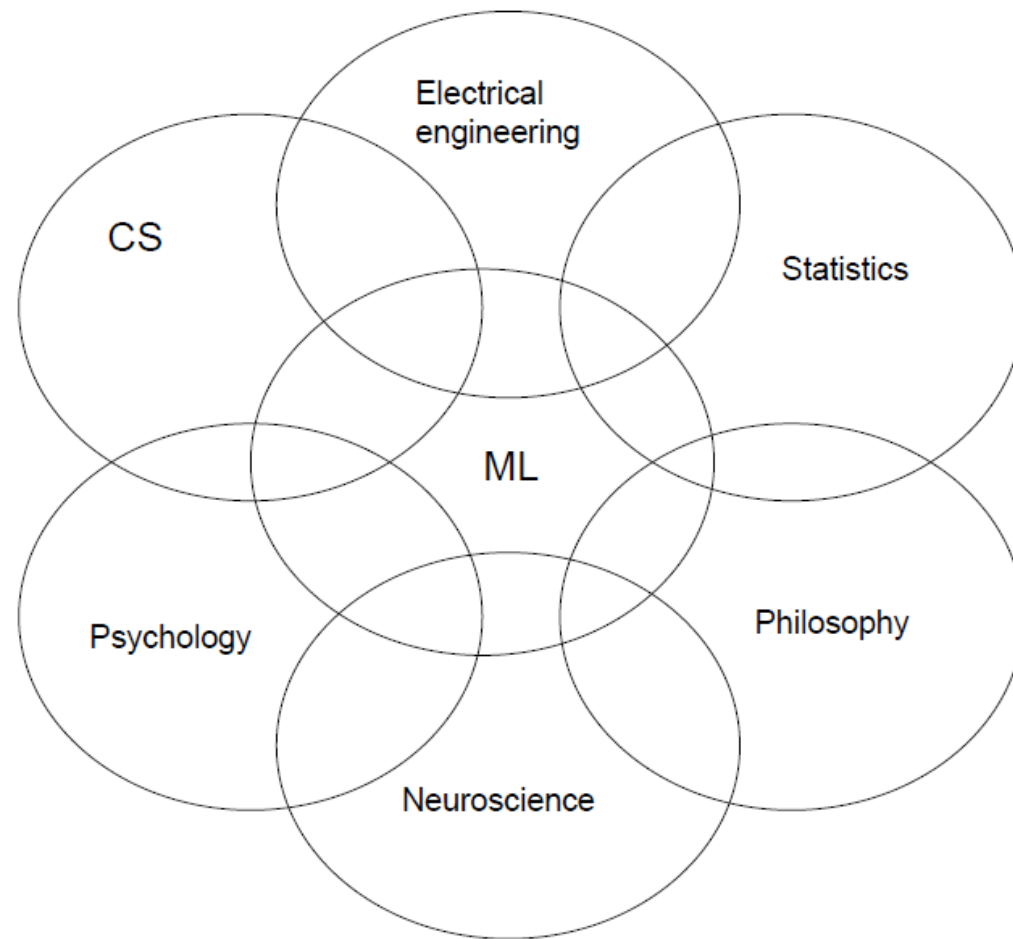
Không phải kì lân biển

Nên chọn quan hệ nào?

Yếu tố thúc đẩy

- Sự phát triển của học máy nhờ vào một số yếu tố sau:
 - Những tiến bộ mới về lý thuyết và thuật toán
 - Dữ liệu trực tuyến tăng trưởng như vũ bão
 - Sức mạnh tính toán đã sẵn sàng
 - Ngành công nghiệp phát triển nở rộ

CÁC LĨNH VỰC LIÊN QUAN

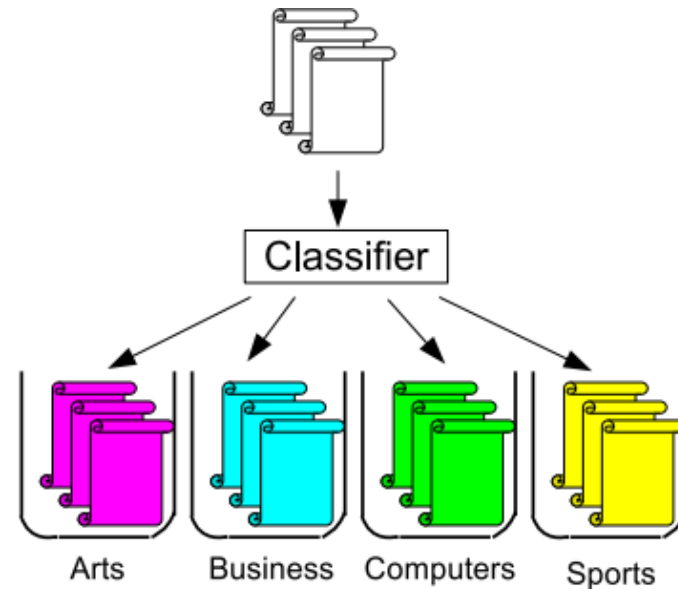


CÁC LĨNH VỰC LIÊN QUAN

- Học máy có liên hệ mật thiết đến lĩnh vực
 - **Xác suất thống kê**: làm khớp mô hình với dữ liệu và kiểm tra kết quả
 - **Khai thác dữ liệu/phân tích dữ liệu thăm dò**: phát hiện hình mẫu trong dữ liệu
 - **Lý thuyết điều khiển thích nghi**: học các mô hình ngay trong quá trình hoạt động và dùng chúng để đạt tới một số mục tiêu
 - **Trí tuệ nhân tạo**: xây dựng cỗ máy thông minh

CÁC DẠNG HỌC MÁY

- **Học có giám sát** (supervised learning)
 - Dựa vào các mẫu huấn luyện đã biết phân lớp.
 - Bộ học được hướng dẫn để nhận diện đặc trưng cho mỗi phân lớp.



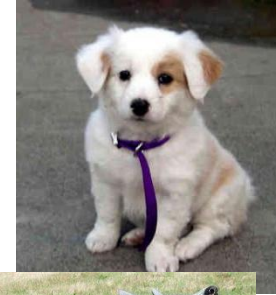
<http://www.cse.lehigh.edu/>

Học có giám sát

Mèo



Chó

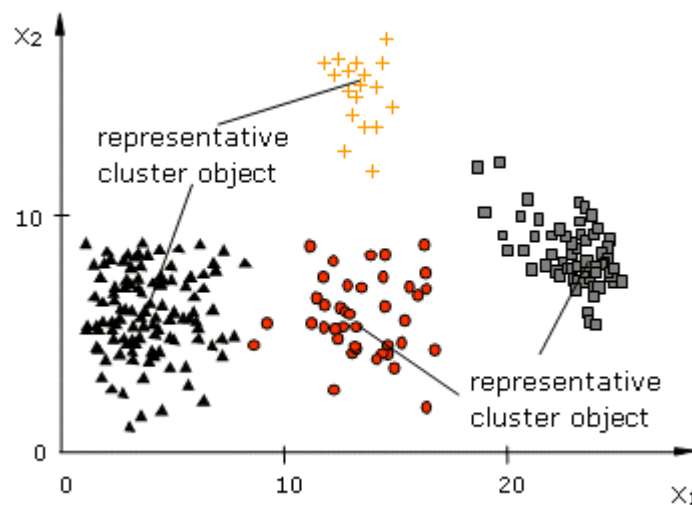


Chuột



CÁC DẠNG HỌC MÁY

- Học không có giám sát (unsupervised learning)
 - Dựa vào mẫu huấn luyện chưa biết phân lớp.
 - Bộ học cần phải tự tìm kiếm thông tin trong hoàn cảnh thiếu sự hướng dẫn.



number of clusters: 4

<http://www.uncertainty-in-engineering.net/>

Học không giám sát

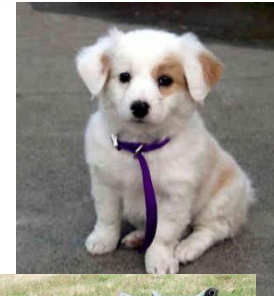


Học có giám sát

Lớp 1



Lớp 2

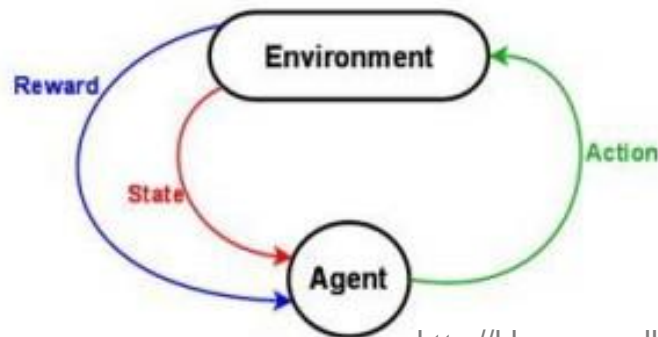


Lớp 3



CÁC DẠNG HỌC MÁY

- **Học tăng cường** (reinforcement learning)
 - Dựa vào mẫu huấn luyện chưa biết phân lớp.
 - Bộ học hành động hướng đến một tình huống và nhận kết quả phản hồi.
 - Tự giải thích phản hồi thông qua việc tạo luật vì không có hướng dẫn.



<http://blogs.cornell.edu/>

Học tăng cường

- **Demo 1:**

Reinforcement Learning Simulation Cart Pole Inverted Pendulum Task

- **Demo 2:**

Evolutionary Reinforcement Learning of Affordances in a NAO Navigation Task Two views ptracking

CÁC MỨC ĐỘ HỌC MÁY

- **Mức độ 1** – Ghi nhớ: dự đoán mẫu đã từng học trước đó.

Mẫu	Quang cảnh	Nhiệt độ	Độ ẩm	Có gió?	Chơi Tennis?
d_1	nắng	85	85	không	không
d_2	nắng	80	90	có	không
d_3	âm u	83	81	không	có
d_4	mưa	70	92	không	có
d_5	âm u	72	90	có	có
$d_?$	nắng	80	90	có	?

CÁC MỨC ĐỘ HỌC MÁY

- **Mức độ 1** – Ghi nhớ: dự đoán mẫu đã từng học trước đó.

Mẫu	Quang cảnh	Nhiệt độ	Độ ẩm	Có gió?	Chơi Tennis?
d ₁	nắng	85	85	không	không
d ₂	nắng	80	90	có	không
d ₃	âm u	83	81	không	có
d ₄	mưa	70	92	không	có
d ₅	âm u	72	90	có	có
d _?	nắng	80	90	có	? (không)

CÁC MỨC ĐỘ HỌC MÁY

- **Mức độ 2** – Lấy trung bình để xử lý nhiễu
 - Dữ liệu giống nhau nhưng kết luận khác nhau

Mẫu	Quang cảnh	Nhiệt độ	Độ ẩm	Có gió?	Chơi Tennis?
d ₁	nắng	85	85	không	không
d ₂	nắng	85	85	không	không
d ₃	nắng	85	85	không	có
d ₄	nắng	85	85	không	có
d ₅	nắng	85	85	không	có
d _?	nắng	85	85	không	?

CÁC MỨC ĐỘ HỌC MÁY

- **Mức độ 2** – Lấy trung bình để xử lý nhiễu
 - Dữ liệu giống nhau nhưng kết luận khác nhau

Mẫu	Quang cảnh	Nhiệt độ	Độ ẩm	Có gió?	Chơi Tennis?
d ₁	nắng	85	85	không	không
d ₂	nắng	85	85	không	không
d ₃	nắng	85	85	không	có
d ₄	nắng	85	85	không	có
d ₅	nắng	85	85	không	có
d _?	nắng	85	85	không	? (có)

CÁC MỨC ĐỘ HỌC MÁY

- **Mức độ 2** – Lấy trung bình để xử lý nhiễu
 - Dữ liệu sai lệch do thiết bị

Mẫu	Quang cảnh	Nhiệt độ	Độ ẩm	Có gió?	Chơi Tennis?
d ₁	nắng	85	85	không	không
d ₂	nắng	80	85	không	không
d ₃	nắng	43	85	không	có
d ₄	nắng	40	85	không	có
d ₅	nắng	42	85	không	có
d _?	nắng	30	85	không	?

CÁC MỨC ĐỘ HỌC MÁY

- **Mức độ 2** – Lấy trung bình để xử lý nhiễu
 - Dữ liệu sai lệch do thiết bị

Mẫu	Quang cảnh	Nhiệt độ	Độ ẩm	Có gió?	Chơi Tennis?
d ₁	nắng	85	85	không	không
d ₂	nắng	80	85	không	không
d ₃	nắng	43	85	không	có
d ₄	nắng	40	85	không	có
d ₅	nắng	42	85	không	có
d _?	nắng	30	85	không	? (có)

CÁC MỨC ĐỘ HỌC MÁY

- **Mức độ 3** – Tổng quát hóa: xử lý dữ liệu chưa từng gặp.

Mẫu	Quang cảnh	Nhiệt độ	Độ ẩm	Có gió?	Chơi Tennis?
d ₁	nắng	85	85	không	không
d ₂	nắng	80	90	có	không
d ₃	âm u	83	8?	không	có
d ₄	mưa	70	92	không	có
d ₅	âm u	72	90	có	có
d _?	mưa	73	93	có	?



PHƯƠNG PHÁP THỐNG KÊ XÁC SUẤT NAÏVE BAYES

ĐỊNH NGHĨA XÁC SUẤT

- Là mức độ tin cậy hay khả năng xảy ra của một sự kiện-một mệnh đề. Ký hiệu $P(A)$.

$$P(A) = \frac{f_A}{N}$$

với N : các kết quả có thể

f_A : số cách mà sự kiện A có thể xảy ra

- Thống kê từ dữ liệu thực
 - Ví dụ: Tung con súc sắc 1000 lần, số lần xuất hiện
 - mặt 1: 162 lần $\Rightarrow P(A=1) = 0.162$
 - mặt 2: 179 lần $\Rightarrow P(A=2) = 0.179$
 - mặt 3: 177 lần $\Rightarrow P(A=3) = 0.177$
 - mặt 4: 172 lần $\Rightarrow P(A=4) = 0.172$
 - mặt 5: 150 lần $\Rightarrow P(A=5) = 0.150$
 - mặt 6: 160 lần $\Rightarrow P(A=6) = 0.160$

CÁCH TÍNH XÁC SUẤT

- Dựa vào mô hình hoặc giá trị lý thuyết
 - Sự kiện: $A = \text{"Ném 1 con súc sắc được mặt số 2"}$
 - $P(A) = 1/6$.

CÚ PHÁP CỦA XÁC SUẤT

- Thành phần cơ bản: **biến ngẫu nhiên**
- Biến ngẫu nhiên **rời rạc**:
 - Ví dụ:
 - Hôm nay trời nắng → Weather có giá trị thuộc <sunny, rainy, cloudy, snow>
- Biến ngẫu nhiên **Boolean**
 - Ví dụ:
 - Có cảnh sát hay không? → Police
 - Có phải là ngày lễ tết không? → Holiday
 - Có đi qua đường chính không? → MainRoad
- Các giá trị trong miền phải **vét cạn** và **loại trừ** lẫn nhau.

CÚ PHÁP CỦA XÁC SUẤT

- Tương tự như logic mệnh đề, các thể hiện có thể có được định nghĩa bằng việc gán giá trị vào biến ngẫu nhiên.
- Mệnh đề cơ sở được xây dựng bằng cách gán một giá trị vào biến ngẫu nhiên.
 - Ví dụ: $Weather = sunny$, $Police = false$ (viết tắt là $\neg Police$)

Xác suất

Phát biểu

- *Khả năng hôm nay có cảnh sát?*

Biểu diễn xác suất

- Police = “Hôm nay có cảnh sát?”
- $P(\text{Police}) = 0.7$
- $\rightarrow P(\neg \text{Police}) = 0.3$ (xác suất hôm nay không có cảnh sát là 0.3)

XÁC SUẤT KẾT HỢP

- Trong thực tế có nhiều yếu tố cùng tồn tại (ngày lễ, thời tiết, đường chính/hẻm) → xác suất kết hợp
- **Phân phối xác suất kết hợp** cho một tập các biến ngẫu nhiên tính xác suất của mỗi sự kiện cơ sở của sự kết hợp các biến ngẫu nhiên này.
 - Ví dụ: Xác suất để ngày hôm nay trời mưa và là ngày lễ là bao nhiêu.
 $P(\text{Weather}=\text{rainy}, \text{Holiday}=\text{true}) = ?$
 - Tổng quát, bảng tất cả các sự kết hợp của hai biến Weather và Holiday

$P(\text{Weather}, \text{Holiday})$ = ma trận 4 cột x 2 dòng

Weather	sunny	rainy	cloudy	snow
Holiday= true	0.144	0.02	0.016	0.02
Holiday=false	0.576	0.08	0.064	0.08

Bài tập

Cho: dữ liệu quan sát hai biến trong 1 năm 365 ngày như sau:

- Số ngày nghỉ: 50 ngày
- Số ngày có cảnh sát: 100 ngày
- Số ngày vừa là ngày nghỉ và có cảnh sát: 30 ngày

Yêu cầu: Hãy lập bảng phân phối xác suất kết hợp giữa hai biến Holiday và Police

	Holiday	\neg Holiday
Police	???	???
\neg Police	???	???

Bài giải

- Số ngày nghỉ: 50 ngày
- Số ngày có cảnh sát: 100 ngày
- Số ngày vừa là ngày nghỉ và có cảnh sát: 30 ngày

$$\rightarrow P(\text{Holiday}, \text{Police}) = 30/365$$

Số ngày nghỉ không có cảnh sát là: $50 - 30 = 20$ ngày

$$\rightarrow P(\text{Holiday}, \neg \text{Police}) = 20/365$$

Số ngày có cảnh sát mà không phải ngày nghỉ: $100 - 30 = 70$

$$\rightarrow P(\neg \text{Holiday}, \text{Police}) = 70/365$$

Số ngày bình thường không có cảnh sát: $(365-50) - 70 = 245$

$$\rightarrow P(\neg \text{Holiday}, \neg \text{Police}) = 245/365$$

Bảng thống kê	Holiday	\neg Holiday
Police	30	70
\neg Police	20	245

Bảng xác suất	Holiday	\neg Holiday
Police	0.0832	0.1918
\neg Police	0.0548	0.6712

Bài giải

- $P(\text{Holiday, Police}) + P(\text{Holiday, } \neg \text{Police}) + P(\neg \text{Holiday, Police}) + P(\neg \text{Holiday, } \neg \text{Police}) = 1$
- $P(\text{Holiday, Police}) + P(\text{Holiday, } \neg \text{Police}) = 0.0832 + 0.0548 = 50/365 = 0.137$
 $P(\text{Holiday}) = 50/365 = 0.137$
 $\rightarrow P(\text{Holiday}) = P(\text{Holiday, Police}) + P(\text{Holiday, } \neg \text{Police})$
- $P(\text{Holiday, } \neg \text{Police}) + P(\neg \text{Holiday, Police}) = 0.0832 + 0.1918 = 0.274$
 $P(\text{Police}) = 100/365 = 0.274$
 $\rightarrow P(\text{Police}) = P(\text{Holiday, } \neg \text{Police}) + P(\neg \text{Holiday, Police})$

Ta có:

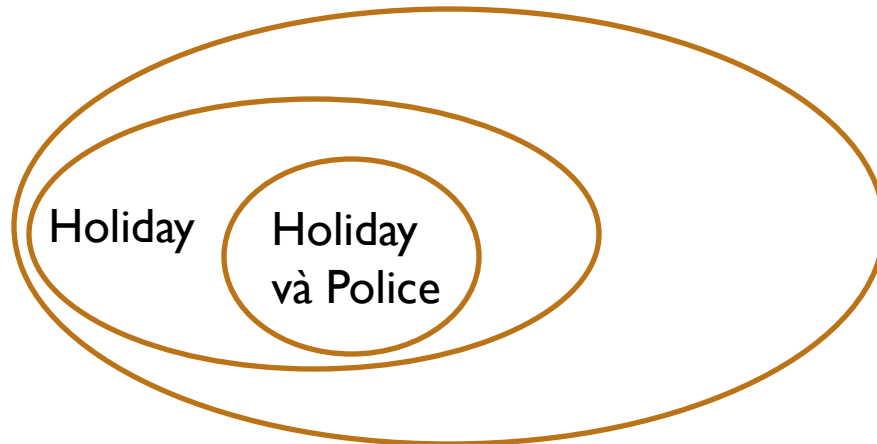
$$P(A) = P(A, B) + P(A, \neg B)$$

Bảng thống kê	Holiday	\neg Holiday
Police	30	70
\neg Police	20	245

Bảng xác suất	Holiday	\neg Holiday
Police	0.0832	0.1918
\neg Police	0.0548	0.6712

XÁC SUẤT ĐIỀU KIỆN

- **Xác suất điều kiện** hay **xác suất hậu nghiệm**: là xác suất của một sự kiện **khi biết trước** một hay nhiều sự kiện khác.
 - Ví dụ ném súc sắc: A: xuất hiện mặt lẻ, B: xuất hiện số 5
 - $P(B) = 1/6$
 - $P(B | A) = 1/3$
 - $P(A | \neg B) = 2/5$
 - Ví dụ: $P(\text{Police} | \text{Holiday})$ tức là đã biết trước hôm nay là holiday.



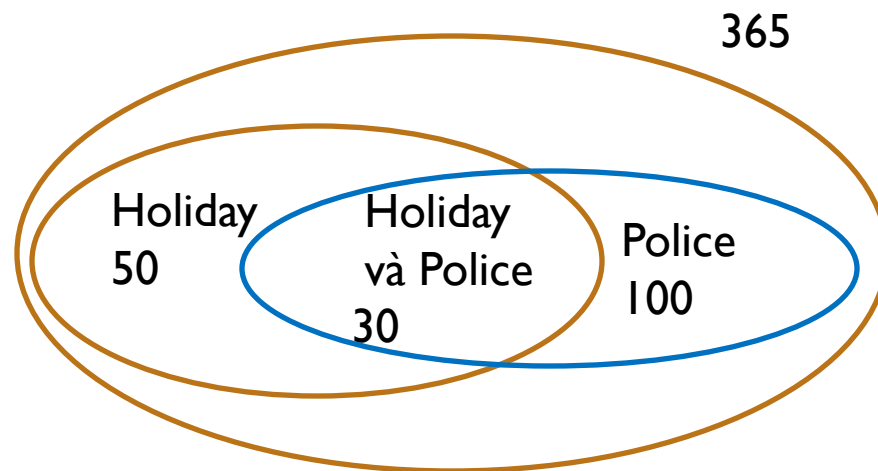
Bài tập

Cho: dữ liệu quan sát hai biết trong 1 năm 365 ngày như sau:

- Số ngày nghỉ: 50 ngày
- Số ngày có cảnh sát: 100 ngày
- Số ngày vừa là ngày nghỉ và có cảnh sát: 30 ngày

Yêu cầu: Tính

- Xác suất có cảnh sát?
- Xác suất có cảnh sát trong những ngày nghỉ?



Bài tập

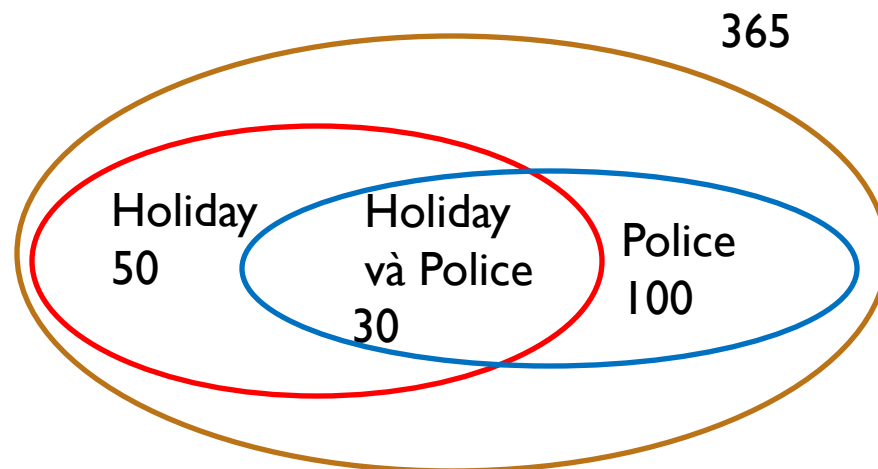
Giải:

Số ngày nghỉ: 50 ngày

- Số ngày có cảnh sát: 100 ngày
- Số ngày vừa là ngày nghỉ và có cảnh sát: 30 ngày

Yêu cầu: Tính

- Xác suất có cảnh sát
- $P(\text{Police}) = 100/365 = 0.274$
- Xác suất có cảnh sát trong những ngày nghỉ
- $P(\text{Police}|\text{Holiday}) = \text{số ngày có Police và Holiday} / \text{số ngày Holiday} = 30/50 = 0.6$



Bài tập

Giải: (tính dùng bảng xác suất)

Số ngày nghỉ: 50 ngày

- Số ngày có cảnh sát: 100 ngày
- Số ngày vừa là ngày nghỉ và có cảnh sát: 30 ngày

Yêu cầu: Tính

- Xác suất có cảnh sát
- $P(\text{Police}) = P(\text{Police}, \text{Holiday}) + P(\text{Police}, \neg\text{Holiday}) = 0.0832 + 0.1918 = 0.274$
- Xác suất có cảnh sát trong những ngày nghỉ
- $P(\text{Police}|\text{Holiday}) = P(\text{Police}, \text{Holiday})/P(\text{Holiday}) = 0.0832/(0.0832+0.0548) = 0.6$

Bảng thống kê	Holiday	\neg Holiday
Police	30	70
\neg Police	20	245

Bảng xác suất	Holiday	\neg Holiday
Police	0.0832	0.1918
\neg Police	0.0548	0.6712

XÁC SUẤT ĐIỀU KIỆN

- Công thức tính xác suất điều kiện

$$P(A | B) = P(A, B) / P(B) \text{ nếu } P(B) > 0$$

- Luật nhân cho ta một công thức khác

$$\begin{aligned} P(A, B) &= P(A | B) P(B) \\ &= P(B | A) P(A) \end{aligned}$$

- Độc lập xác suất: A và B là hai sự kiện độc lập (sự kiện này xảy ra hay không hoàn toàn không ảnh hưởng đến sự kiện kia)

$$P(B | A) = P(B)$$

khi đó

$$P(A, B) = P(A) P(B)$$

PHÂN LỚP NAÏVE BAYES

- Gọi X_1, X_2, \dots, X_n là các thuộc tính trong tập dữ liệu, Y là thuộc tính phân lớp. Giả sử X_1, X_2, \dots, X_n độc lập điều kiện với nhau khi cho Y .
- Thuật toán dựa trên luật suy diễn xác suất Bayes

$$P(Y | X) = \frac{P(X | Y)P(Y)}{P(X)}$$

- Cập nhật xác suất của giả thiết dựa trên chứng cứ.
 - Chọn giả thiết có xác suất lớn nhất sau khi tích hợp các chứng cứ.
- Naïve Bayes đặc biệt hữu ích cho các bài toán có **nhieu** đặc trưng.

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
?	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
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
?	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

- 1. $P(\text{Trượt tuyết} = \text{có}) = ?$
- 2. $P(\text{Trượt tuyết} = \text{không}) = ?$
- 3. $P(\text{Tuyết} = \text{băng} \mid \text{Trượt tuyết} = \text{có}) = ?$
 $P(\text{Tuyết} = \text{ẩm} \mid \text{Trượt tuyết} = \text{có}) = ?$
 $P(\text{Tuyết} = \text{khô} \mid \text{Trượt tuyết} = \text{có}) = ?$
- 4. $P(\text{Tuyết} = \text{băng} \mid \text{Trượt tuyết} = \text{không}) = ?$
 $P(\text{Tuyết} = \text{ẩm} \mid \text{Trượt tuyết} = \text{không}) = ?$
 $P(\text{Tuyết} = \text{khô} \mid \text{Trượt tuyết} = \text{không}) = ?$

BÀI TẬP VÍ DỤ

- 1. $P(\text{Trượt tuyết} = \text{có}) = ?$
- 2. $P(\text{Trượt tuyết} = \text{không}) = ?$
- 3. $P(\text{Tuyết} = \text{băng} \mid \text{Trượt tuyết} = \text{có}) = ?$
 $P(\text{Tuyết} = \text{ẩm} \mid \text{Trượt tuyết} = \text{có}) = ?$
 $P(\text{Tuyết} = \text{khô} \mid \text{Trượt tuyết} = \text{có}) = ?$
- 4. $P(\text{Tuyết} = \text{băng} \mid \text{Trượt tuyết} = \text{không}) = ?$
 $P(\text{Tuyết} = \text{ẩm} \mid \text{Trượt tuyết} = \text{không}) = ?$
 $P(\text{Tuyết} = \text{khô} \mid \text{Trượt tuyết} = \text{không}) = ?$
- 5. $P(\text{Thời tiết} = \text{gió} \mid \text{Trượt tuyết} = \text{có}) = ?$
 $P(\text{Thời tiết} = \text{nắng} \mid \text{Trượt tuyết} = \text{có}) = ?$
 $P(\text{Thời tiết} = \text{sương mù} \mid \text{Trượt tuyết} = \text{có}) = ?$
- 6. $P(\text{Thời tiết} = \text{gió} \mid \text{Trượt tuyết} = \text{không}) = ?$
 $P(\text{Thời tiết} = \text{nắng} \mid \text{Trượt tuyết} = \text{không}) = ?$
 $P(\text{Thời tiết} = \text{sương mù} \mid \text{Trượt tuyết} = \text{không}) = ?$

BÀI TẬP VÍ DỤ - ĐÁP ÁN

- 1. $P(\text{Trượt tuyết} = \text{có}) = 9/14$
- 2. $P(\text{Trượt tuyết} = \text{không}) = 5/14$
- 3. $P(\text{Tuyết} = \text{băng} \mid \text{Trượt tuyết} = \text{có}) = 0/9$
 $P(\text{Tuyết} = \text{ẩm} \mid \text{Trượt tuyết} = \text{có}) = 2/9$
 $P(\text{Tuyết} = \text{khô} \mid \text{Trượt tuyết} = \text{có}) = 7/9$
- 4. $P(\text{Tuyết} = \text{băng} \mid \text{Trượt tuyết} = \text{không}) = 3/5$
 $P(\text{Tuyết} = \text{ẩm} \mid \text{Trượt tuyết} = \text{không}) = 1/5$
 $P(\text{Tuyết} = \text{khô} \mid \text{Trượt tuyết} = \text{không}) = 1/5$
- 5. $P(\text{Thời tiết} = \text{gió} \mid \text{Trượt tuyết} = \text{có}) = 2/9$
 $P(\text{Thời tiết} = \text{nắng} \mid \text{Trượt tuyết} = \text{có}) = 5/9$
 $P(\text{Thời tiết} = \text{sương mù} \mid \text{Trượt tuyết} = \text{có}) = 2/9$
- 6. $P(\text{Thời tiết} = \text{gió} \mid \text{Trượt tuyết} = \text{không}) = 1/5$
 $P(\text{Thời tiết} = \text{nắng} \mid \text{Trượt tuyết} = \text{không}) = 1/5$
 $P(\text{Thời tiết} = \text{sương mù} \mid \text{Trượt tuyết} = \text{không}) = 3/5$

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
?	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

•

$P(\text{Mùa} = \text{vắng} \mid \text{Trượt tuyết} = \text{có}) = ?$

$P(\text{Mùa} = \text{vừa phải} \mid \text{Trượt tuyết} = \text{có}) = ?$

• 2. $P(\text{Mùa} = \text{cao điểm} \mid \text{Trượt tuyết} = \text{không}) = ?$

$P(\text{Mùa} = \text{vắng} \mid \text{Trượt tuyết} = \text{không}) = ?$

$P(\text{Mùa} = \text{vừa phải} \mid \text{Trượt tuyết} = \text{không}) = ?$

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
?	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

- 3. $P(\text{Sức khỏe} = \text{bị thương} \mid \text{Trượt tuyết} = \text{có}) = ?$
 $P(\text{Sức khỏe} = \text{mệt mỏi} \mid \text{Trượt tuyết} = \text{có}) = ?$
 $P(\text{Sức khỏe} = \text{tốt} \mid \text{Trượt tuyết} = \text{có}) = ?$
- 4. $P(\text{Sức khỏe} = \text{bị thương} \mid \text{Trượt tuyết} = \text{không}) = ?$
 $P(\text{Sức khỏe} = \text{mệt mỏi} \mid \text{Trượt tuyết} = \text{không}) = ?$
 $P(\text{Sức khỏe} = \text{tốt} \mid \text{Trượt tuyết} = \text{không}) = ?$

BÀI TẬP VÍ DỤ - ĐÁP ÁN

- 1. $P(\text{Mùa} = \text{cao điểm} \mid \text{Trượt tuyết} = \text{có}) = 1/9$
 $P(\text{Mùa} = \text{vắng} \mid \text{Trượt tuyết} = \text{có}) = 6/9$
 $P(\text{Mùa} = \text{vừa phải} \mid \text{Trượt tuyết} = \text{có}) = 2/9$
- 2. $P(\text{Mùa} = \text{cao điểm} \mid \text{Trượt tuyết} = \text{không}) = 1/5$
 $P(\text{Mùa} = \text{vắng} \mid \text{Trượt tuyết} = \text{không}) = 3/5$
 $P(\text{Mùa} = \text{vừa phải} \mid \text{Trượt tuyết} = \text{không}) = 1/5$
- 3. $P(\text{Sức khỏe} = \text{bị thương} \mid \text{Trượt tuyết} = \text{có}) = 0/9$
 $P(\text{Sức khỏe} = \text{mệt mỏi} \mid \text{Trượt tuyết} = \text{có}) = 0/9$
 $P(\text{Sức khỏe} = \text{tốt} \mid \text{Trượt tuyết} = \text{có}) = 9/9$
- 4. $P(\text{Sức khỏe} = \text{bị thương} \mid \text{Trượt tuyết} = \text{không}) = 2/5$
 $P(\text{Sức khỏe} = \text{mệt mỏi} \mid \text{Trượt tuyết} = \text{không}) = 1/5$
 $P(\text{Sức khỏe} = \text{tốt} \mid \text{Trượt tuyết} = \text{không}) = 2/5$

DỰ ĐOÁN MẪU MỚI

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
15	ảm	nắng	vắng	mệt mỏi	?

- $P(\text{Trượt tuyết} = \text{có} \mid X)$
$$= \frac{P(X \mid \text{Trượt tuyết} = \text{có}) \times P(\text{Trượt tuyết} = \text{có})}{P(X)}$$
- $P(X \mid \text{Trượt tuyết} = \text{có}) =$
 - $P(\text{Tuyết} = \text{ảm} \mid \text{Trượt tuyết} = \text{có})$
 - $\times P(\text{Thời tiết} = \text{nắng} \mid \text{Trượt tuyết} = \text{có})$
 - $\times P(\text{Mùa} = \text{vắng} \mid \text{Trượt tuyết} = \text{có})$
 - $\times P(\text{Sức khỏe} = \text{mệt mỏi} \mid \text{Trượt tuyết} = \text{có})$

DỰ ĐOÁN MẪU MỚI

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
15	ẩm	nắng	vắng	mệt mỏi	?

X

- $P(\text{Trượt tuyết} = \text{có} \mid X)$
= $P(\text{Tuyết} = \text{ẩm} \mid \text{Trượt tuyết} = \text{có})$
× $P(\text{Thời tiết} = \text{nắng} \mid \text{Trượt tuyết} = \text{có})$
× $P(\text{Mùa} = \text{vắng} \mid \text{Trượt tuyết} = \text{có})$
× $P(\text{Sức khỏe} = \text{mệt mỏi} \mid \text{Trượt tuyết} = \text{có})$
× $P(\text{Trượt tuyết} = \text{có}) / P(X)$
- Tính tương tự cho $P(\text{Trượt tuyết} = \text{không} \mid X)$

DỰ ĐOÁN MẪU MỚI

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
15	ảm	nắng	vắng	mệt mỏi	?

- Gọi $P(\text{Trượt tuyết} = \text{có} \mid X) = S1$ và $P(\text{Trượt tuyết} = \text{không} \mid X) = S2$
 - Nếu $S1 < S2$ thì **Không** đi trượt tuyết
 - Nếu $S1 > S2$ thì **Có** đi trượt tuyết
 - Nếu $S1 = S2$ thì chọn tùy ý **Có** hoặc **Không**
- Do $S1$ và $S2$ có mẫu số giống nhau ($=P(X)$) nên chỉ cần tính tử số \rightarrow Bỏ $P(X)$

BÀI TẬP VÍ DỤ

- Hãy dự đoán kết luận cho mẫu bên dưới.

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
15	ẩm	nắng	vắng	mệt mỏi	?

BÀI TẬP VÍ DỤ - ĐÁP ÁN

- Dự đoán kết luận cho mẫu bên dưới.

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
15	ảm	nắng	vắng	mệt mỏi	?

- $P(\text{Trượt tuyết} = \text{có} \mid X)$
 $= 2/9 * 5/9 * 6/9 * 0/9 * 9/14 = 0$
- $P(\text{Trượt tuyết} = \text{không} \mid X)$
 $= 1/5 * 1/5 * 3/5 * 1/5 * 5/14 > 0$
- Vậy kết luận là **Không đi trượt tuyết.**

BÀI TẬP VÍ DỤ - ĐÁP ÁN

- Dự đoán kết luận cho mẫu bên dưới.

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
15	ảm	nắng	vắng	mệt mỏi	?

- $P(\text{Trượt tuyết} = \text{có} \mid X)$
 $= 2/9 * 5/9 * 6/9 * 0/9 * 9/14 = 0 \quad (1)$
- $P(\text{Trượt tuyết} = \text{không} \mid X)$
 $= 1/5 * 1/5 * 3/5 * 1/5 * 5/14 > 0 \quad (2)$
- Vậy kết luận là **Không đi trượt tuyết**.
Tuy nhiên, chú ý rằng các xác suất của (1) trừ 0/9 đều > (2)

LÀM TRƠN LAPLACE

- Được áp dụng để khắc phục trường hợp $P(X_i = u \mid Y = v) = 0$ do không có mẫu nào trong dữ liệu huấn luyện thỏa mãn điều kiện.

$$P(Y = v) = \frac{N(Y = v) + 1}{N_{DL} + m}$$

$$P(X_i = u \mid Y = v) = \frac{N(X_i = u \wedge Y = v) + 1}{N(Y = v) + r}$$

- Với m là số phân lớp, r là số giá trị của thuộc tính X_i

BÀI TẬP VÍ DỤ

- Hãy áp dụng làm trơn trên tất cả các giá trị xác suất và dự đoán lại kết luận cho mẫu bên dưới.

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
15	ấm	nắng	vắng	mệt mỏi	?

BÀI TẬP VÍ DỤ - LÀM TRƠN

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
?	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

- $P(\text{Trượt tuyết} = \text{có}) = 9/14 \Rightarrow ?$
 $P(\text{Trượt tuyết} = \text{không}) = 5/14 \Rightarrow ?$
- $P(\text{Tuyết} = \text{băng} \mid \text{Trượt tuyết} = \text{có}) = 0/9 \Rightarrow ?$
 $P(\text{Tuyết} = \text{ẩm} \mid \text{Trượt tuyết} = \text{có}) = 2/9 \Rightarrow ?$
 $P(\text{Tuyết} = \text{khô} \mid \text{Trượt tuyết} = \text{có}) = 7/9 \Rightarrow ?$
- $P(\text{Tuyết} = \text{băng} \mid \text{Trượt tuyết} = \text{không}) = 3/5 \Rightarrow ?$
 $P(\text{Tuyết} = \text{ẩm} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow ?$
 $P(\text{Tuyết} = \text{khô} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow ?$

BÀI TẬP VÍ DỤ - LÀM TRƠN

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
?	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

- $P(\text{Thời tiết} = \text{gió} \mid \text{Trượt tuyết} = \text{có}) = 2/9 \Rightarrow ?$
 $P(\text{Thời tiết} = \text{nắng} \mid \text{Trượt tuyết} = \text{có}) = 5/9 \Rightarrow ?$
 $P(\text{Thời tiết} = \text{sương mù} \mid \text{Trượt tuyết} = \text{có}) = 2/9 \Rightarrow ?$
- $P(\text{Thời tiết} = \text{gió} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow ?$
 $P(\text{Thời tiết} = \text{nắng} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow ?$
 $P(\text{Thời tiết} = \text{sương mù} \mid \text{Trượt tuyết} = \text{không}) = 3/5 \Rightarrow ?$

BÀI TẬP VÍ DỤ - LÀM TRƠN ĐÁP ÁN

- $P(\text{Trượt tuyết} = \text{có}) = 9/14 \Rightarrow 10/16$
 $P(\text{Trượt tuyết} = \text{không}) = 5/14 \Rightarrow 6/16$
- $P(\text{Tuyết} = \text{băng} \mid \text{Trượt tuyết} = \text{có}) = 0/9 \Rightarrow 1/12$
 $P(\text{Tuyết} = \text{ẩm} \mid \text{Trượt tuyết} = \text{có}) = 2/9 \Rightarrow 3/12$
 $P(\text{Tuyết} = \text{khô} \mid \text{Trượt tuyết} = \text{có}) = 7/9 \Rightarrow 8/12$
- $P(\text{Tuyết} = \text{băng} \mid \text{Trượt tuyết} = \text{không}) = 3/5 \Rightarrow 4/8$
 $P(\text{Tuyết} = \text{ẩm} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow 2/8$
 $P(\text{Tuyết} = \text{khô} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow 2/8$
- $P(\text{Thời tiết} = \text{gió} \mid \text{Trượt tuyết} = \text{có}) = 2/9 \Rightarrow 3/12$
 $P(\text{Thời tiết} = \text{nắng} \mid \text{Trượt tuyết} = \text{có}) = 5/9 \Rightarrow 6/12$
 $P(\text{Thời tiết} = \text{sương mù} \mid \text{Trượt tuyết} = \text{có}) = 2/9 \Rightarrow 3/12$
- $P(\text{Thời tiết} = \text{gió} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow 2/8$
 $P(\text{Thời tiết} = \text{nắng} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow 2/8$
 $P(\text{Thời tiết} = \text{sương mù} \mid \text{Trượt tuyết} = \text{không}) = 3/5 \Rightarrow 4/8$

BÀI TẬP VÍ DỤ

- $P(\text{Mùa} = \text{cao điểm} \mid \text{Trượt tuyết} = \text{có}) = 1/9 \Rightarrow ?$
 $P(\text{Mùa} = \text{vắng} \mid \text{Trượt tuyết} = \text{có}) = 6/9 \Rightarrow ?$
 $P(\text{Mùa} = \text{vừa phải} \mid \text{Trượt tuyết} = \text{có}) = 2/9 \Rightarrow ?$
- $P(\text{Mùa} = \text{cao điểm} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow ?$
 $P(\text{Mùa} = \text{vắng} \mid \text{Trượt tuyết} = \text{không}) = 3/5 \Rightarrow ?$
 $P(\text{Mùa} = \text{vừa phải} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow ?$
- $P(\text{Sức khỏe} = \text{bị thương} \mid \text{Trượt tuyết} = \text{có}) = 0/9 \Rightarrow ?$
 $P(\text{Sức khỏe} = \text{mệt mỏi} \mid \text{Trượt tuyết} = \text{có}) = 0/9 \Rightarrow ?$
 $P(\text{Sức khỏe} = \text{tốt} \mid \text{Trượt tuyết} = \text{có}) = 9/9 \Rightarrow ?$
- $P(\text{Sức khỏe} = \text{bị thương} \mid \text{Trượt tuyết} = \text{không}) = 2/5 \Rightarrow ?$
 $P(\text{Sức khỏe} = \text{mệt mỏi} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow ?$
 $P(\text{Sức khỏe} = \text{tốt} \mid \text{Trượt tuyết} = \text{không}) = 2/5 \Rightarrow ?$

BÀI TẬP VÍ DỤ - ĐÁP ÁN

- $P(\text{Mùa} = \text{cao điểm} \mid \text{Trượt tuyết} = \text{có}) = 1/9 \Rightarrow 2/12$
 $P(\text{Mùa} = \text{vắng} \mid \text{Trượt tuyết} = \text{có}) = 6/9 \Rightarrow 7/12$
 $P(\text{Mùa} = \text{vừa phải} \mid \text{Trượt tuyết} = \text{có}) = 2/9 \Rightarrow 3/12$
- $P(\text{Mùa} = \text{cao điểm} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow 2/8$
 $P(\text{Mùa} = \text{vắng} \mid \text{Trượt tuyết} = \text{không}) = 3/5 \Rightarrow 4/8$
 $P(\text{Mùa} = \text{vừa phải} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow 2/8$
- $P(\text{Sức khỏe} = \text{bị thương} \mid \text{Trượt tuyết} = \text{có}) = 0/9 \Rightarrow 1/12$
 $P(\text{Sức khỏe} = \text{mệt mỏi} \mid \text{Trượt tuyết} = \text{có}) = 0/9 \Rightarrow 1/12$
 $P(\text{Sức khỏe} = \text{tốt} \mid \text{Trượt tuyết} = \text{có}) = 9/9 \Rightarrow 10/12$
- $P(\text{Sức khỏe} = \text{bị thương} \mid \text{Trượt tuyết} = \text{không}) = 2/5 \Rightarrow 3/8$
 $P(\text{Sức khỏe} = \text{mệt mỏi} \mid \text{Trượt tuyết} = \text{không}) = 1/5 \Rightarrow 2/8$
 $P(\text{Sức khỏe} = \text{tốt} \mid \text{Trượt tuyết} = \text{không}) = 2/5 \Rightarrow 3/8$

BÀI TẬP VÍ DỤ - ĐÁP ÁN

- Dự đoán kết luận cho mẫu bên dưới.

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
15	ẩm	nắng	vắng	mệt mỏi	?

- $P(\text{Trượt tuyết} = \text{có} \mid X)$
 $= 3/12 * 6/12 * 7/12 * 1/12 * 10/16 = 0.0038$
- $P(\text{Trượt tuyết} = \text{không} \mid X)$
 $= 2/8 * 2/8 * 4/8 * 2/8 * 6/16 = 0.0029$
- Vậy kết luận là **Đi trượt tuyết**.

XỬ LÝ DỮ LIỆU THIỂU

- Nếu mẫu sắp dự đoán bị thiếu giá trị tại một hay một số thuộc tính thì không xét các thuộc tính đó vào tích xác suất.

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
16	-	gió	vừa phải	tốt	?

- $P(\text{Trượt tuyết} = \text{có} \mid X)$
= $P(\text{Trượt tuyết} = \text{có})$
× $P(\text{Thời tiết} = \text{gió} \mid \text{Trượt tuyết} = \text{có})$
× $P(\text{Mùa} = \text{vừa phải} \mid \text{Trượt tuyết} = \text{có})$
× $P(\text{Sức khỏe} = \text{tốt} \mid \text{Trượt tuyết} = \text{có})$
- Tính tương tự cho $P(\text{Trượt tuyết} = \text{không} \mid X)$

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