



Khoa Công Nghệ Thông Tin
Trường Đại Học Cần Thơ



Máy học tự động Machine learning

Đỗ Thanh Nghi
dtnghi@cit.ctu.edu.vn

Cần Thơ
12-02-2019

Nội dung

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Nội dung

- **Tại sao máy học là cần thiết?**
- Máy học và các bài toán
- Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Máy học tự động

■ Tại sao máy học là cần thiết?

- Hiện tượng vật lý, sinh học, tài chính, v.v.: quá phức tạp để được mô tả một cách xác định
- Kỹ thuật học máy: hỗ trợ / thay thế con người
- Giải quyết vấn đề trong các lĩnh vực khoa học
- Phân tích dữ liệu
- Nhận dạng khuôn mặt, dấu vân tay
- Nhận dạng tiếng nói, tổng hợp tiếng nói
- Điều khiển robot, xe tự hành
- Trợ lý ảo Chatbot
- Phân lớp văn bản, tóm tắt văn bản
- Trò chơi, v.v.

Nội dung

- Tại sao máy học là cần thiết?
- **Máy học và các bài toán**
- Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Máy học tự động là gì?

■ Wikipedia

- **Học** là tập hợp các cơ chế nhằm tiếp thu cái mới hoặc bổ sung, trau dồi các kiến thức, kỹ năng, kinh nghiệm, giá trị, nhận thức
- **Máy học tự động** đề cập đến các khái niệm, phân tích thiết kế, phát triển, cài đặt các phương pháp cho phép máy tính tiến hóa thông qua một quy trình có hệ thống, nhằm thực hiện các tác vụ khó hoặc không thể thực hiện bằng phương pháp thuật toán cổ điển hơn

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Máy học tự động là gì?

■ Máy học tự động

- Liên quan đến khái niệm, thiết kế và phát triển chương trình bất kỳ để cải thiện phương thức hoạt động của nó bằng cách tiếp thu kiến thức và kỹ năng mới
- Máy tính sử dụng các thuật toán cho phép phát triển tiến hóa các hành vi, dữ liệu trừu tượng và các mẫu
- Ngành khoa học thúc đẩy máy tính có thể hành động/ứng xử mà không cần lập trình riêng biệt
- Thuật toán học từ những sai lầm để phát triển kết quả tốt nhất mà không cần sự can thiệp của con người
- Nhằm mục đích thiết kế các chương trình có thể tự động cải thiện kinh nghiệm

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Máy học tự động là gì?

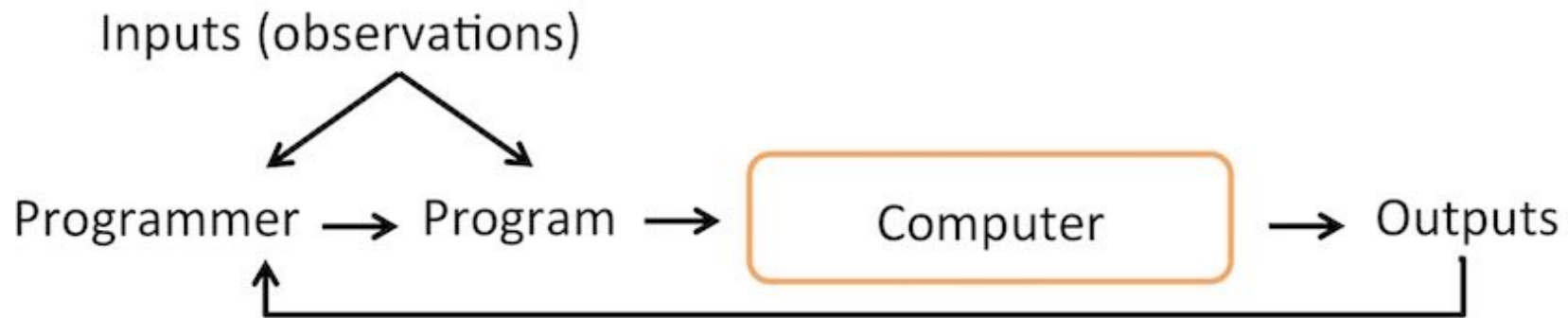
■ Máy học tự động

- Liên quan đến việc nghiên cứu và xây dựng các kỹ thuật cho phép máy tính học tự động từ dữ liệu để giải quyết những vấn đề cụ thể
- Một chương trình máy tính được xem là học cách thực thi một lớp nhiệm vụ thông qua trải nghiệm, đối với thang đo năng lực nếu như dùng thang đo năng lực ta đo thấy năng lực thực thi của chương trình có tiến bộ sau khi đã trải nghiệm

Lập trình truyền thống

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

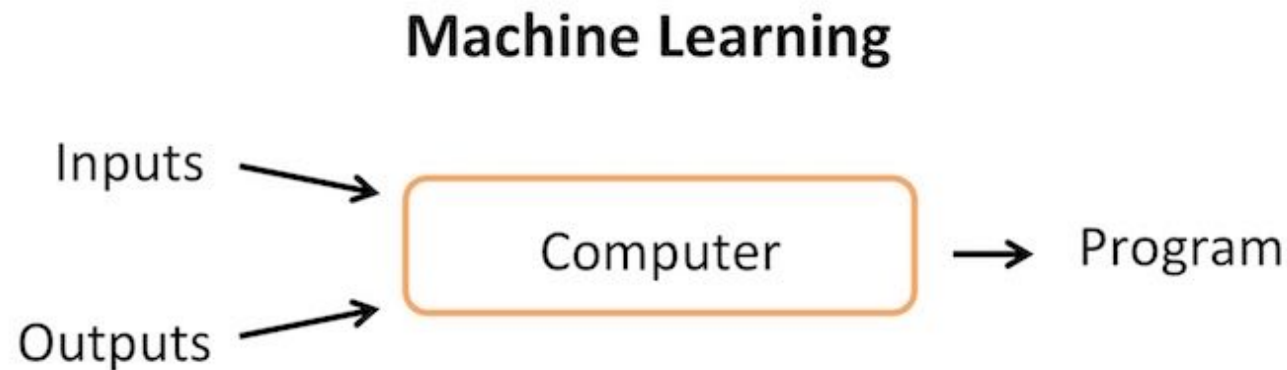
The Traditional Programming Paradigm



- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Máy học tự động

Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed
– Arthur Samuel (1959)



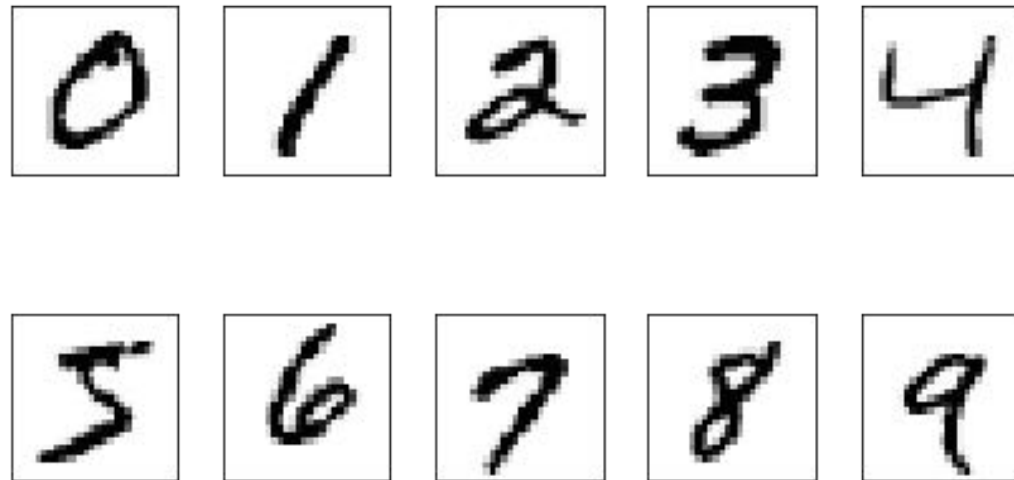
- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Máy học tự động

“A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .”

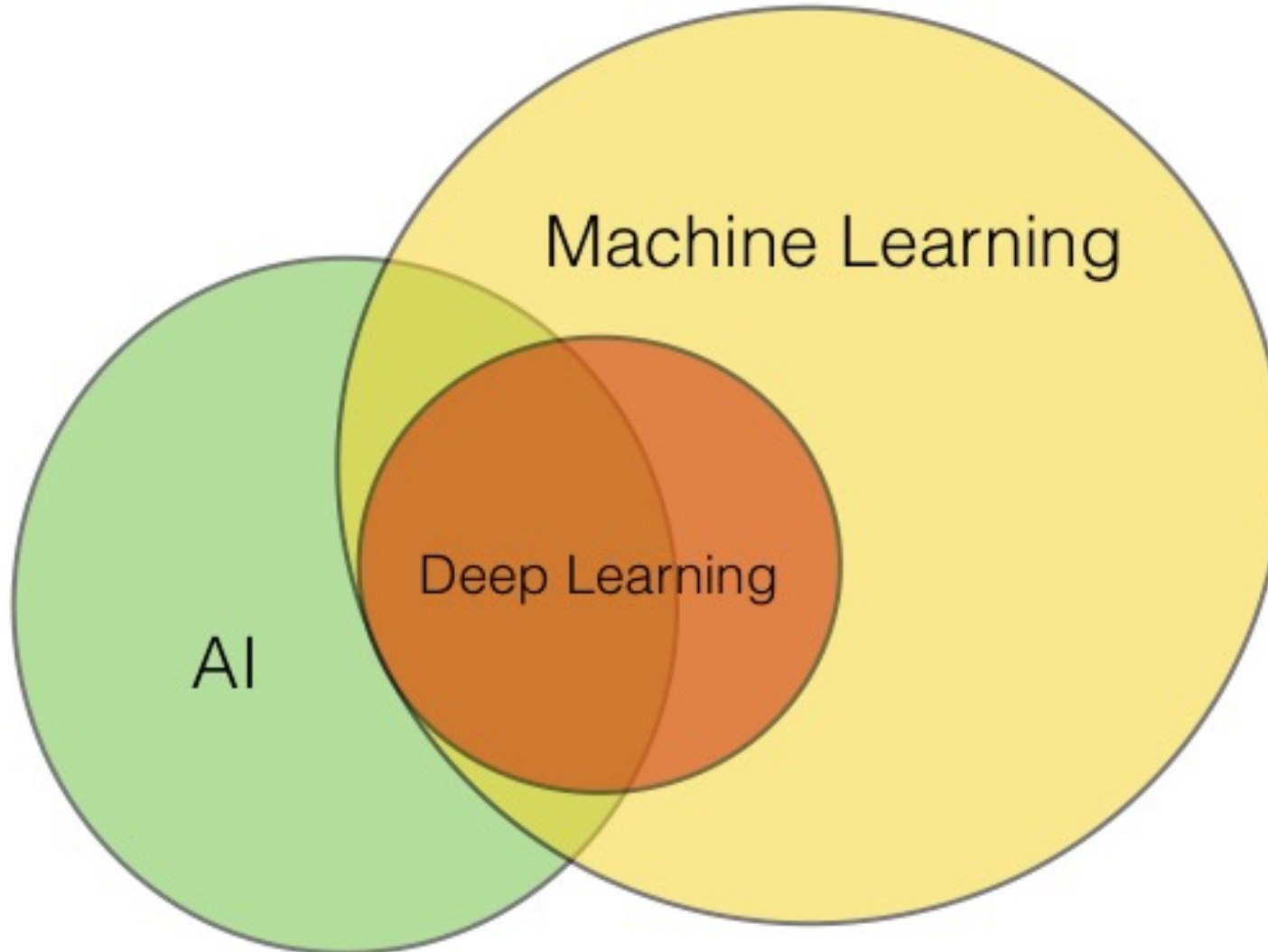
— Tom Mitchell, Professor at Carnegie Mellon University

Handwriting Recognition Example:



- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Machine Learning, AI, Deep Learning



Các bài toán

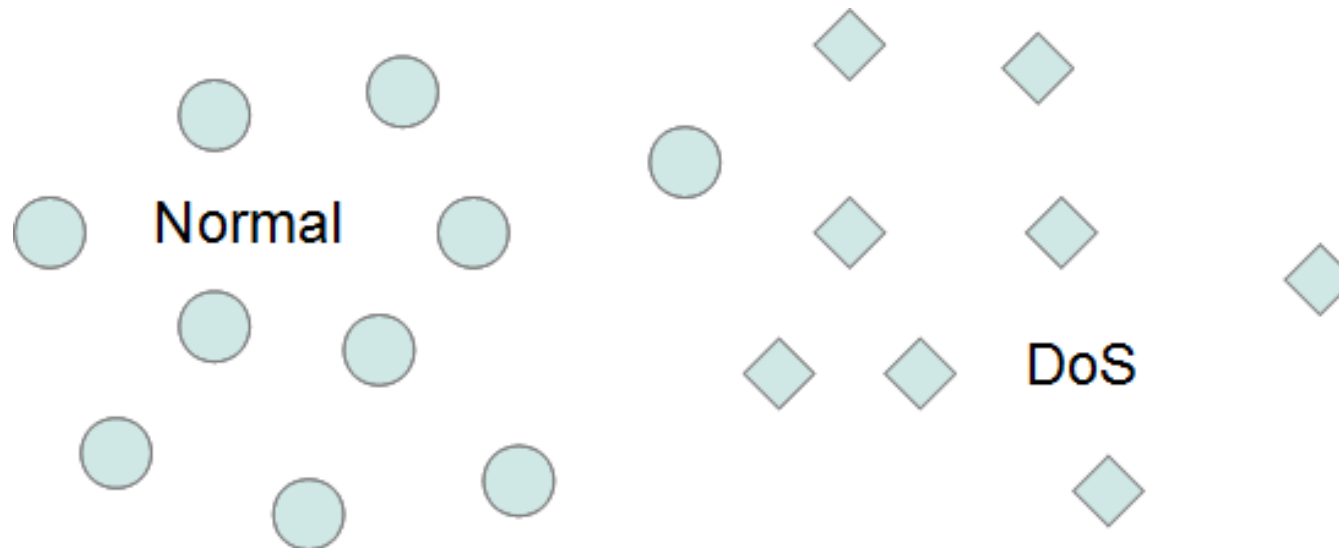
■ Học có giám sát

- Biến phụ thuộc Y cần dự đoán, được mô tả bởi m phần tử mà chúng ta biết n biến dự đoán độc lập được tổng hợp trong X
- Tập huấn luyện $D_{Train} = \{(X_1, Y_1), \dots, (X_m, Y_m)\}$
- Giả sử có tập huấn luyện D_{Train} , chúng ta cần tìm hàm ϕ của n biến dự đoán sao cho biến cần dự đoán Y được dự đoán tốt nhất bởi hàm ϕ : $Y = \phi(X) + \varepsilon$
- Học được giám sát bởi nhãn (lớp, giá trị dự đoán) của từng phần tử của tập dữ liệu có m phần tử: những giá trị Y_i
- Phương pháp: mô hình hồi quy, cây quyết định, mạng nơ-ron nhân tạo, máy học véc-tơ hỗ trợ SVM, k NN, etc

Các bài toán

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

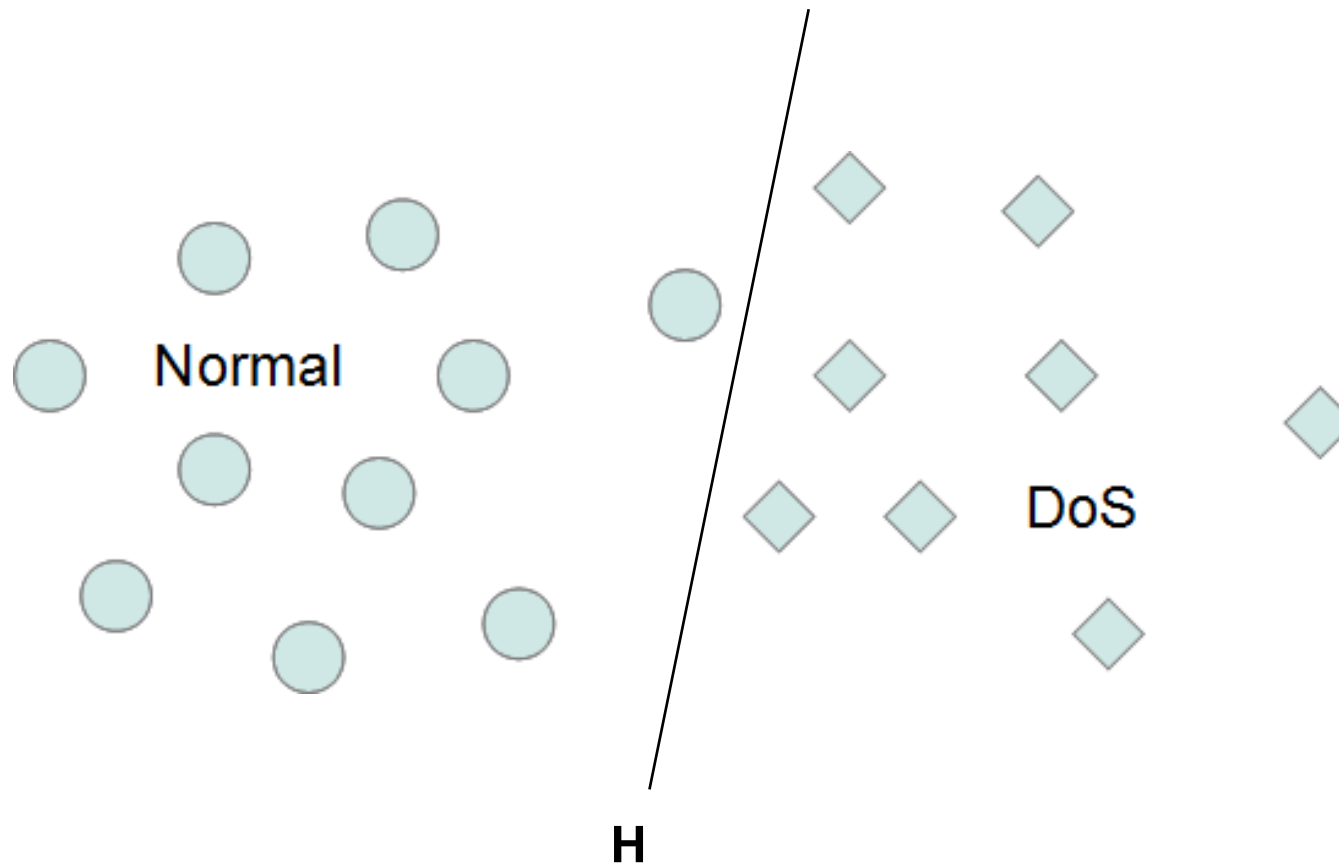
■ Học có giám sát: **phân lớp - classification**



Các bài toán

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

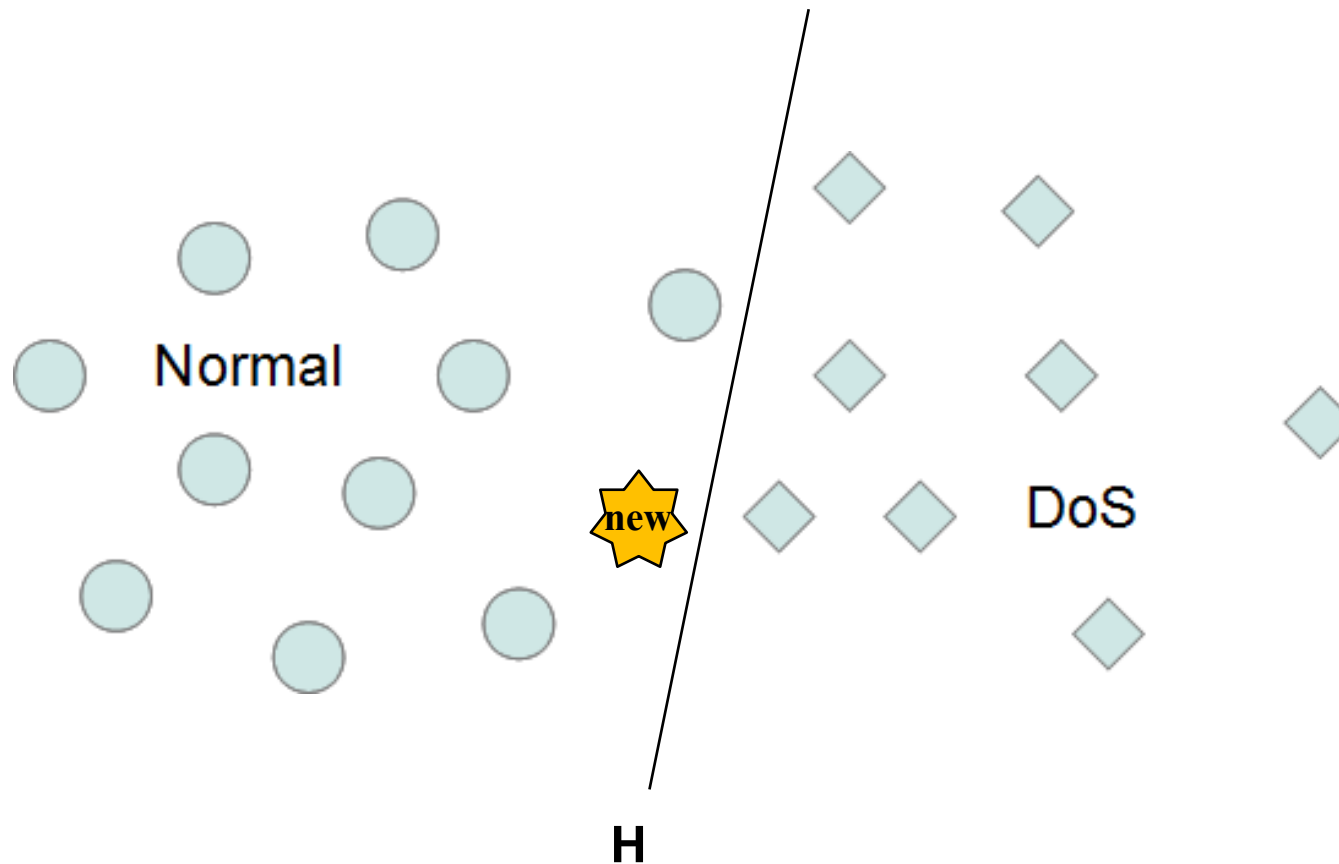
■ Học có giám sát: **phân lớp - classification**



Các bài toán

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

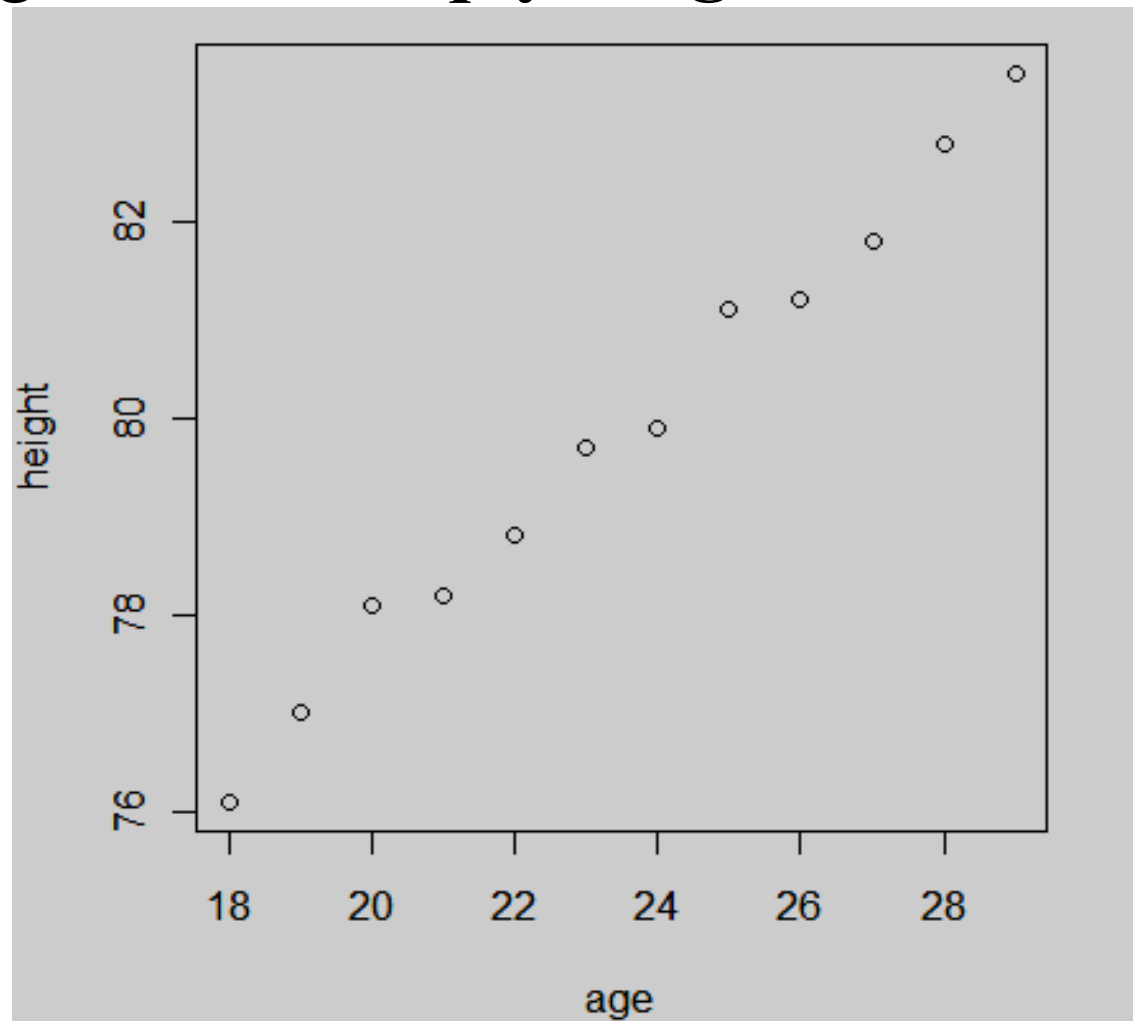
■ Học có giám sát: **phân lớp - classification**



Các bài toán

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

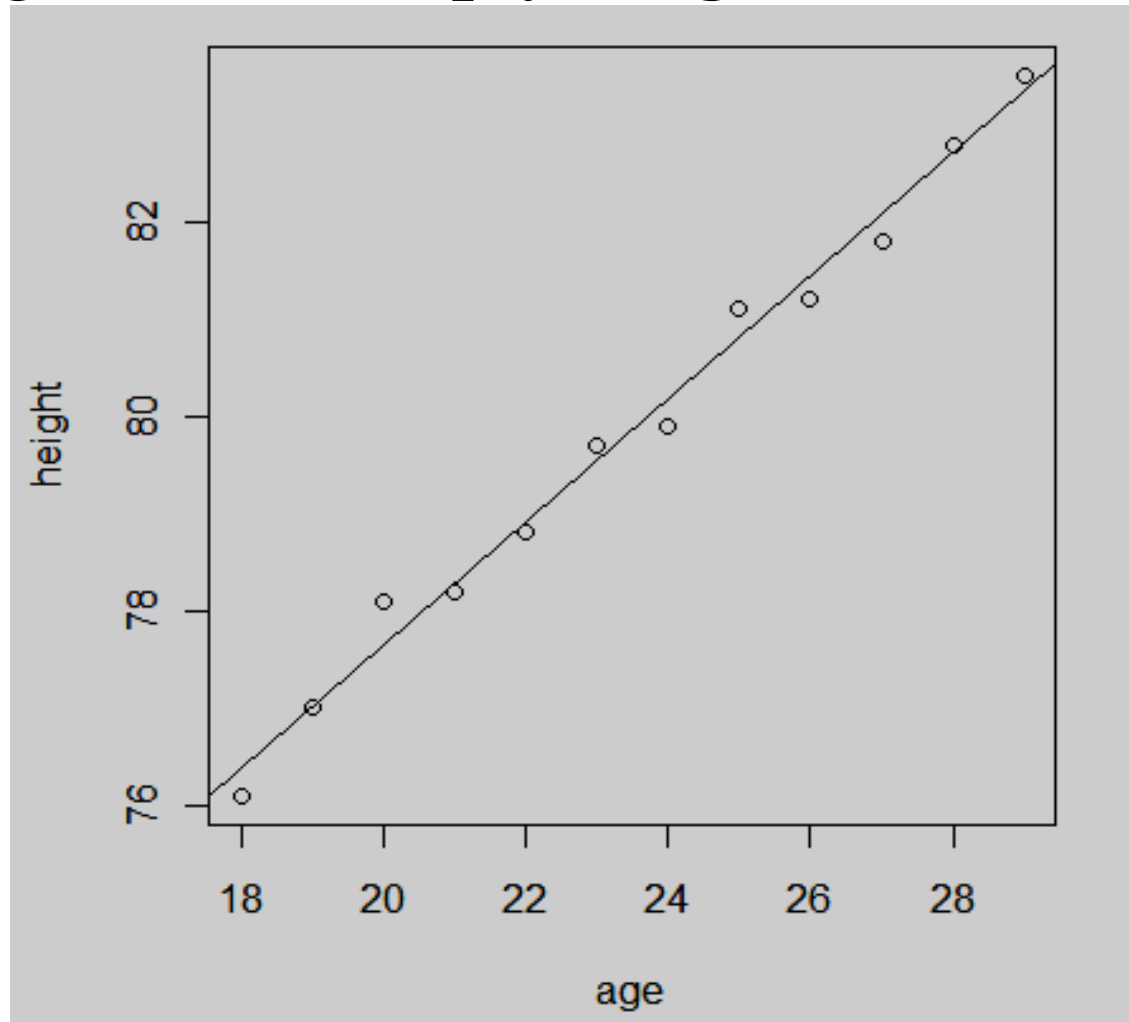
■ Học có giám sát: **hồi quy - regression**



Các bài toán

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

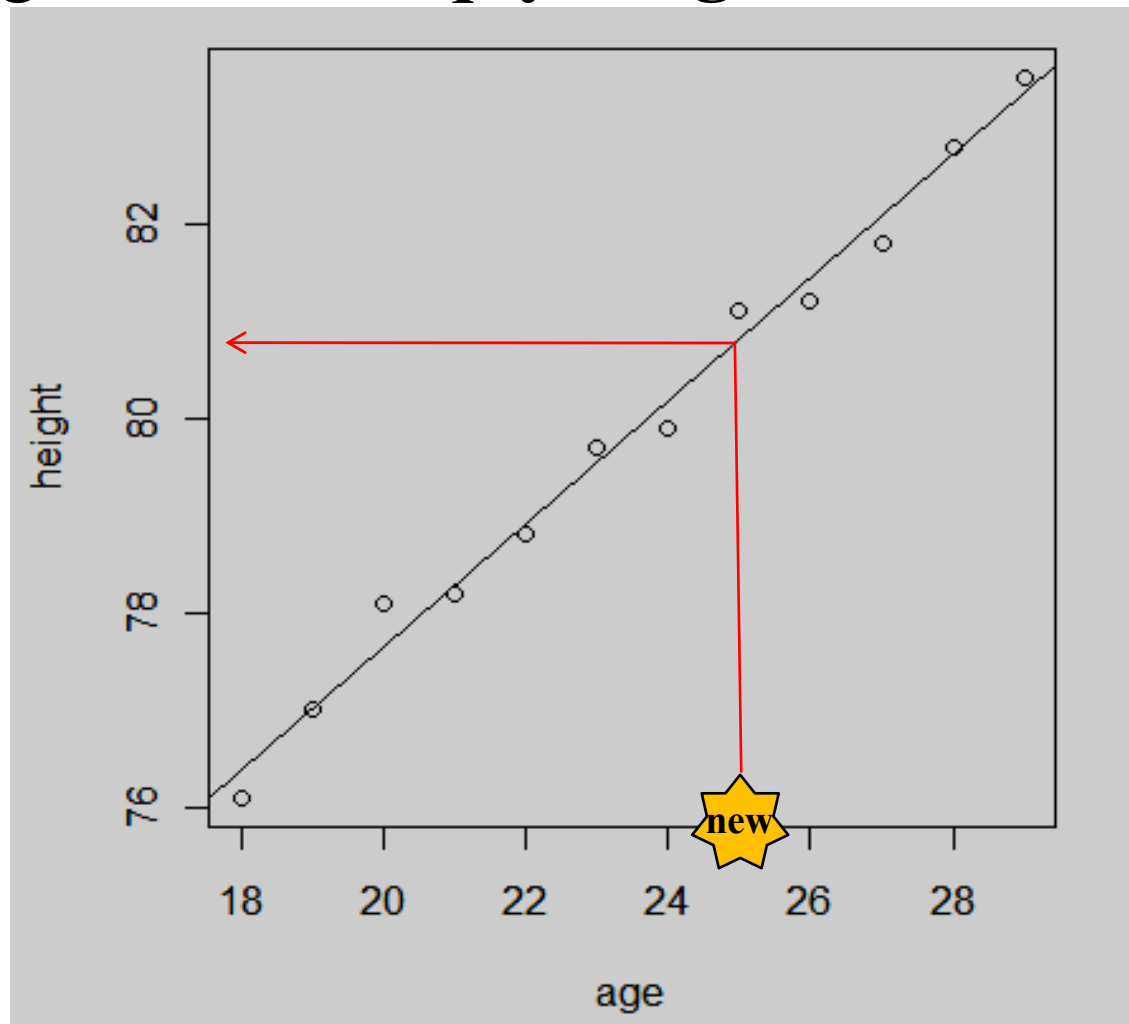
■ Học có giám sát: **hồi quy - regression**



Các bài toán

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

■ Học có giám sát: **hồi quy - regression**



- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Các bài toán

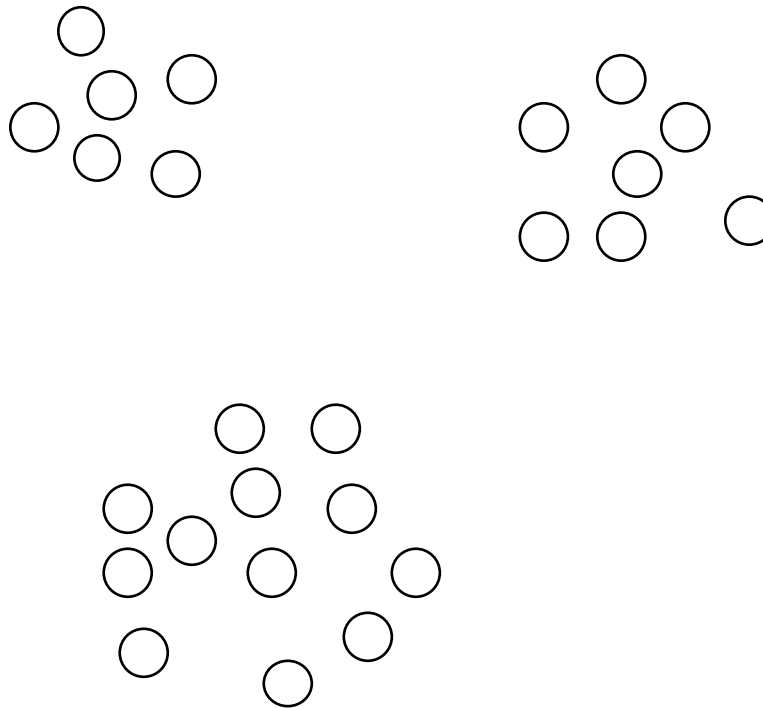
■ Học không giám sát

- Không có biến Y cần dự đoán, nhưng có m phần tử được mô tả bởi n biến dự đoán độc lập được tổng hợp trong X
- Tập huấn luyện $D_{Train} = \{X_1, \dots, X_m\}$
- Nhằm mục đích gom nhóm (cụm, nhóm các phần tử có đặc điểm chung) trong dữ liệu
- Phương pháp: gom nhóm phân cấp, k -means, etc

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Các bài toán

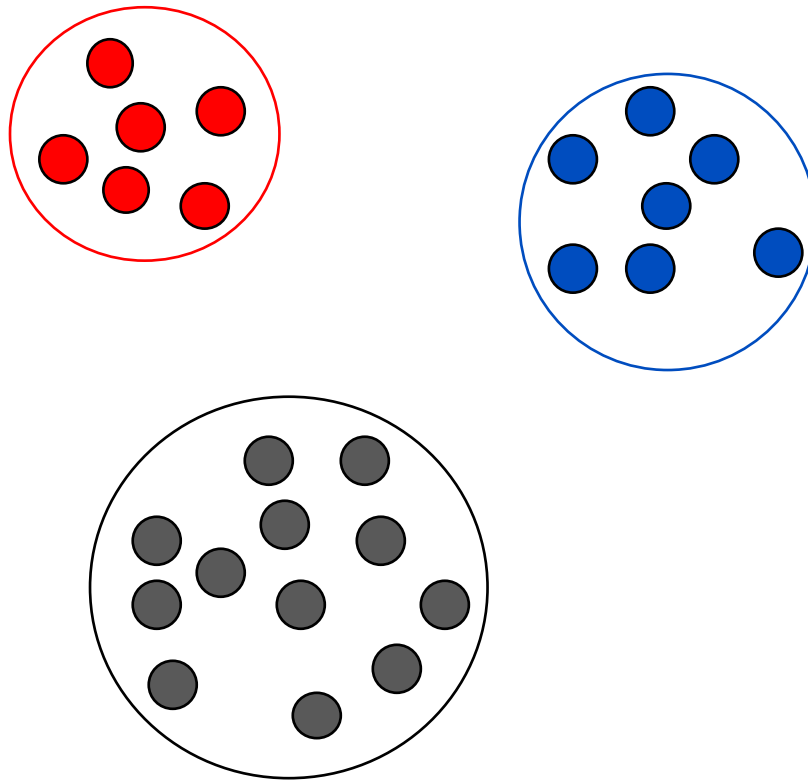
■ Học không giám sát: **gom nhóm - clustering**



Các bài toán

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

■ Học không giám sát: **gom nhóm - clustering**



- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Các vấn đề

■ Dữ liệu

- m, n : hợp lý; Giả định về mô hình là tuyến tính và phân phối dữ liệu được xác định; Mẫu dữ liệu được cho tuân theo phân phối (Gauss, nhị thức, Poisson); Các mô hình phân tích thống kê có thể xử lý tốt
- Vấn đề trở nên khó hơn để xử lý nếu mẫu dữ liệu hạn chế
- Giả định về phân phối dữ liệu có thể chưa biết
- Mô hình dự đoán phi tuyến
- Kích cỡ m, n dữ liệu lớn

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Các vấn đề

■ Lựa chọn phương pháp

- Nhiều phương pháp máy học tự động được đề xuất
- Không có phương pháp nào là tốt nhất cho tất cả các bài toán, các lĩnh vực ứng dụng
- Một phương pháp có ưu điểm và khuyết điểm cho một vấn đề đặt ra trong thực tiễn, phụ thuộc vào bản chất của dữ liệu, năng lực của lớp hàm ϕ cần học để ước lượng Y
- Cần thiết phải biết cách so sánh các phương pháp để chọn phương pháp phù hợp nhất
- Tiêu chí so sánh có thể dựa trên ước tính lỗi (hồi quy hoặc phân lớp)

Nội dung

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- **Ứng dụng**

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Ứng dụng

- Sinh-tin học
- Phân loại trình tự DNA
- Hỗ trợ chẩn đoán
- Thị giác máy tính
- Công nhận chữ viết tay
- Nhận dạng giọng nói
- Phát hiện gian lận
- Phân tích thị trường tài chính

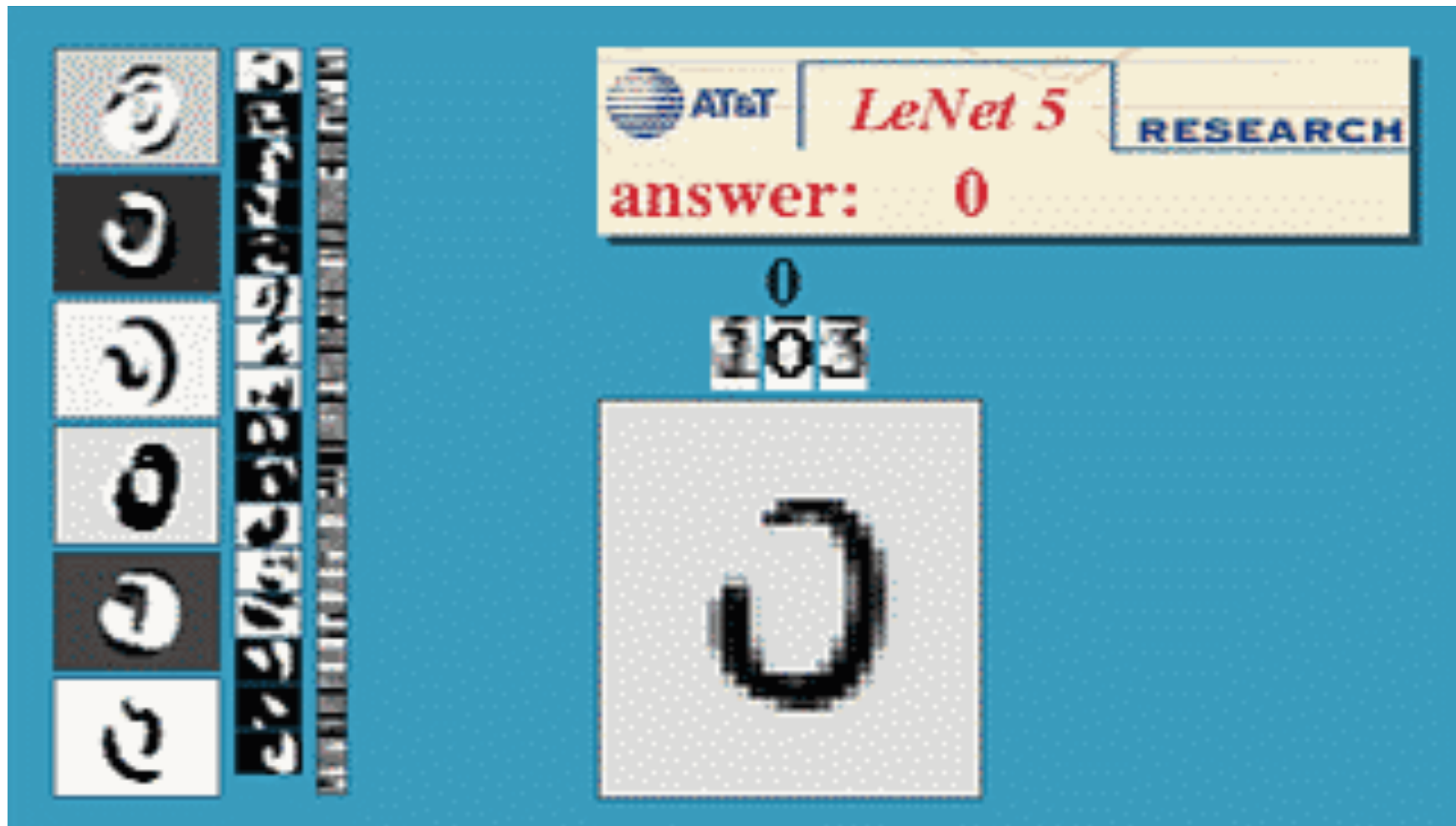
- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Ứng dụng

- Tìm kiếm thông tin
- Phân tích và lập chỉ mục hình ảnh và video
- Tìm kiếm hình ảnh theo nội dung
- Trò chơi (hệ thống tự học)
- Xử lý ngôn ngữ tự nhiên
- Hệ thống gợi ý
- Phân tích quan điểm tự động
- Điều khiển robot
- Xe tự hành

Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng



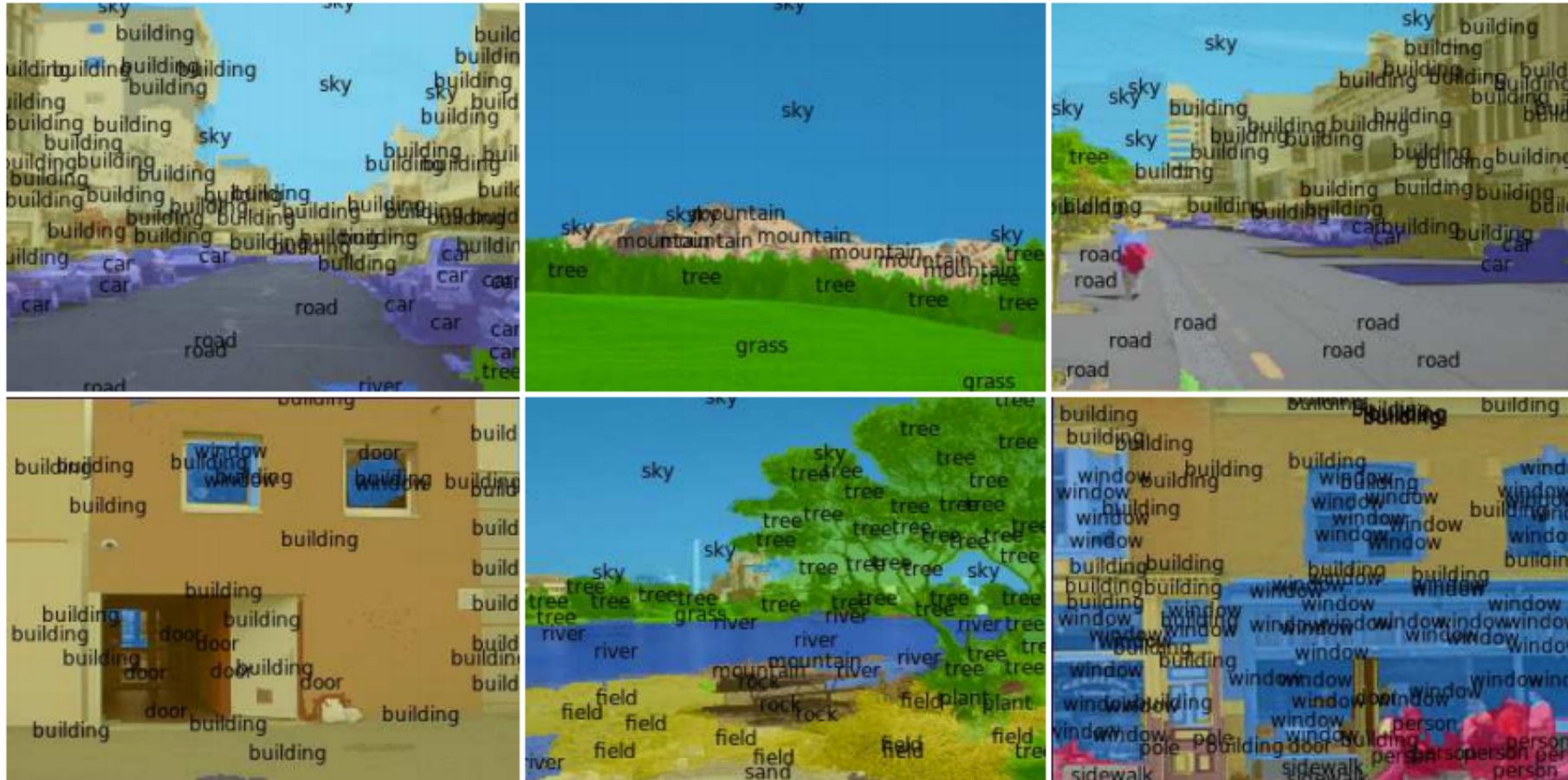
Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng



Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng



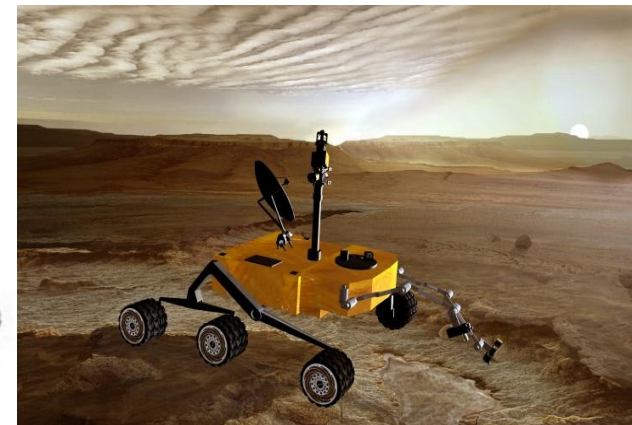
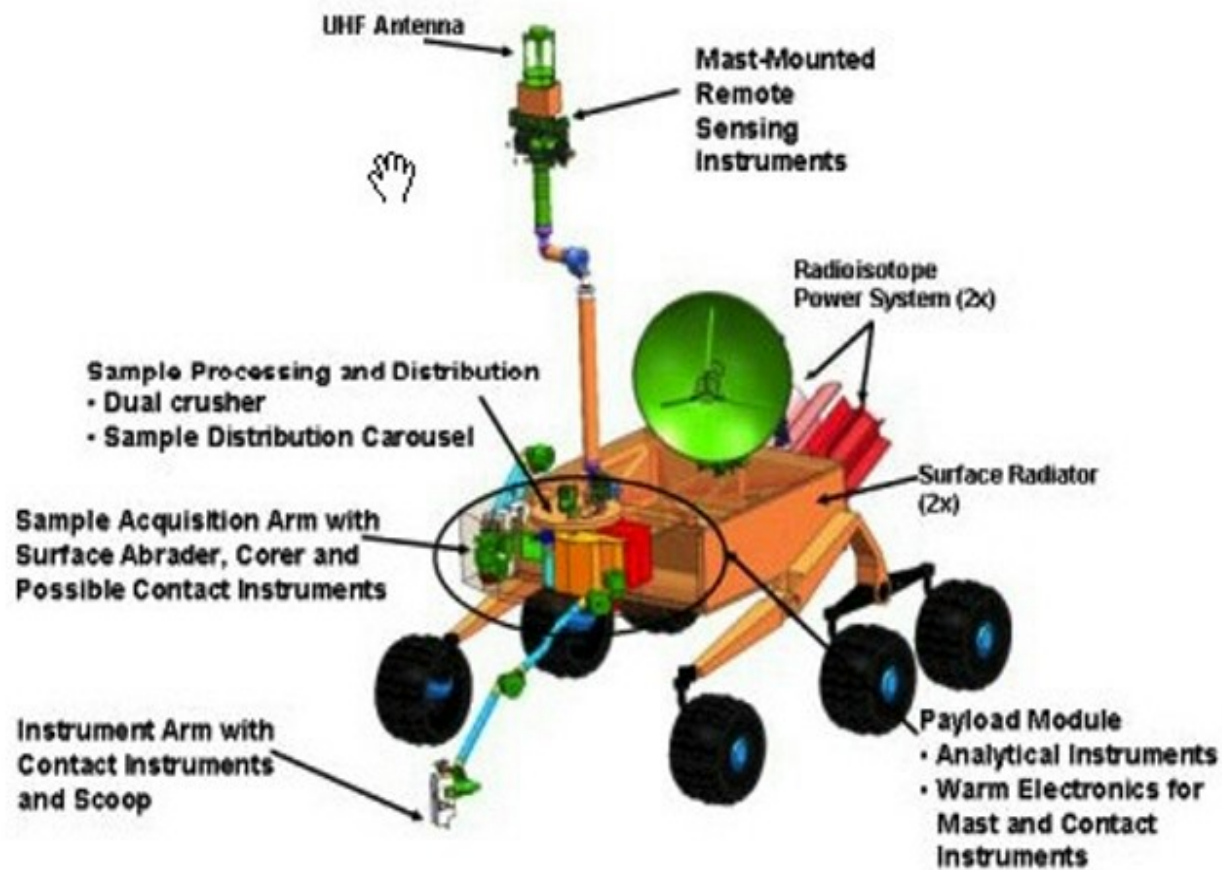
Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng



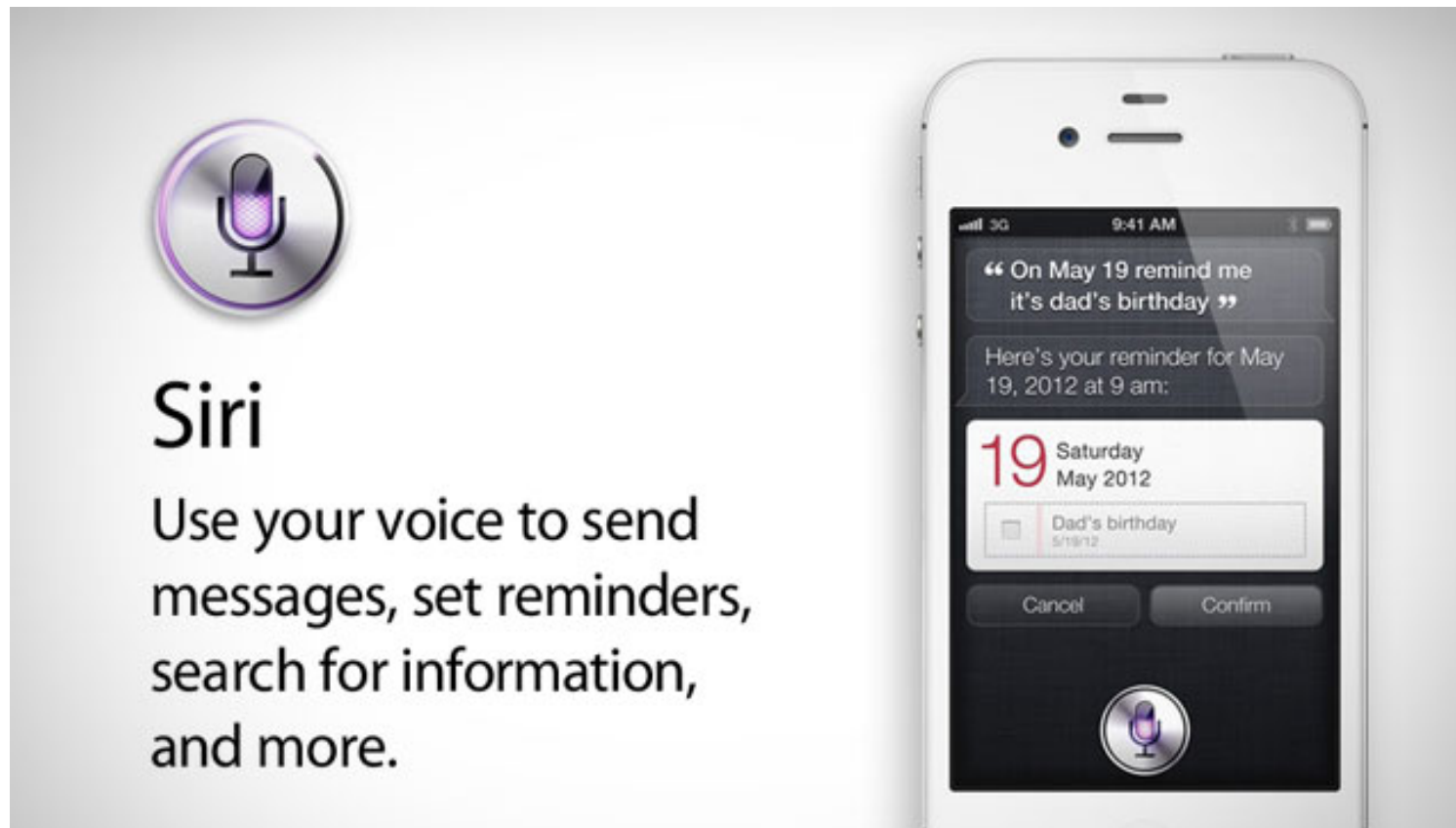
Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng



Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng



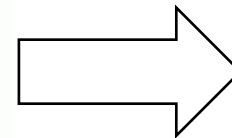
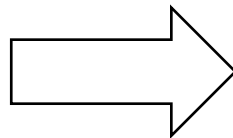
Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

From: xor@clotho.som.rpi.edu
(Joe Schwartz) Subject: Re:
NUTEK FACES APPLE'S
WRATH (article!!!!) READ

In article <davea-
120493231310@129.228.20.182>
davea@xetron.com (David P.
Alverson) writes:

I believe Apple has a patent on the
region features of QuickDraw. A
mac clone would have to
implement regions. This is why
Apple's comment was that they
believe it is not possible to make a
Mac clone without infringing on
their patents. They may have
other patents like this.



Category: **Comp.mac.**

Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng

Topics

gene	0.04
dna	0.02
genetic	0.01
...	

life	0.02
evolve	0.01
organism	0.01
...	

brain	0.04
neuron	0.02
nerve	0.01
...	

data	0.02
number	0.02
computer	0.01
...	

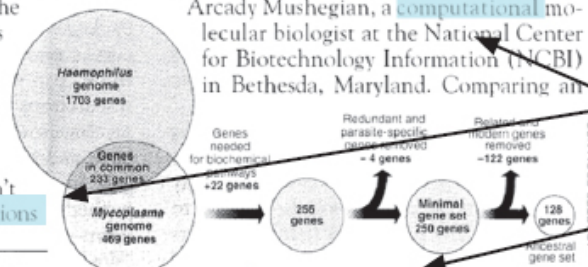
Documents

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions

"are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

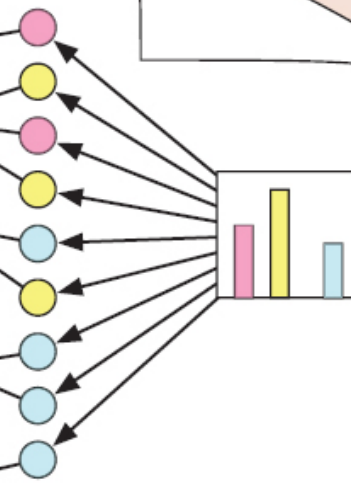


* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

SCIENCE • VOL. 272 • 24 MAY 1996

Topic proportions and assignments



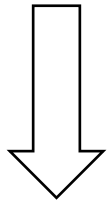
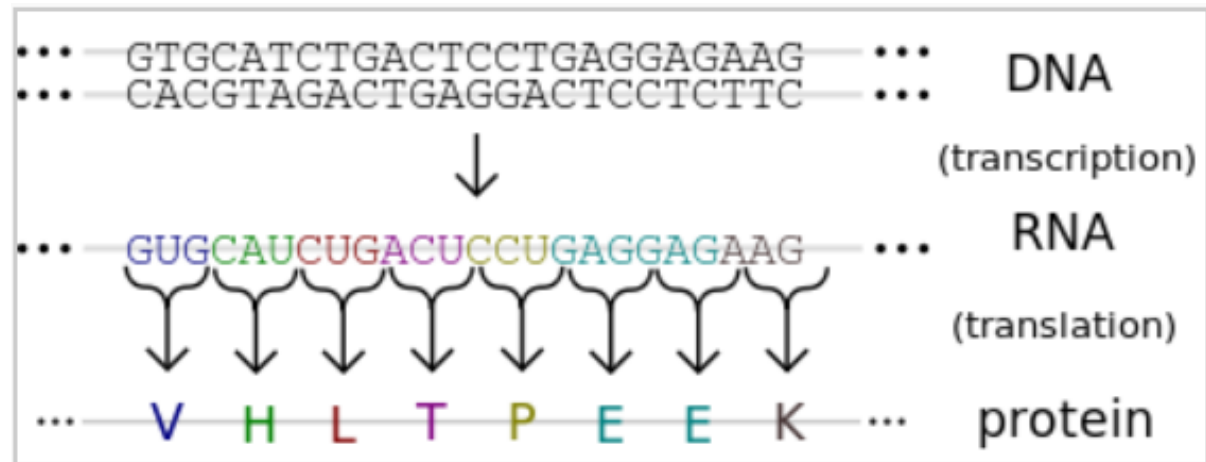
Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng



Ứng dụng

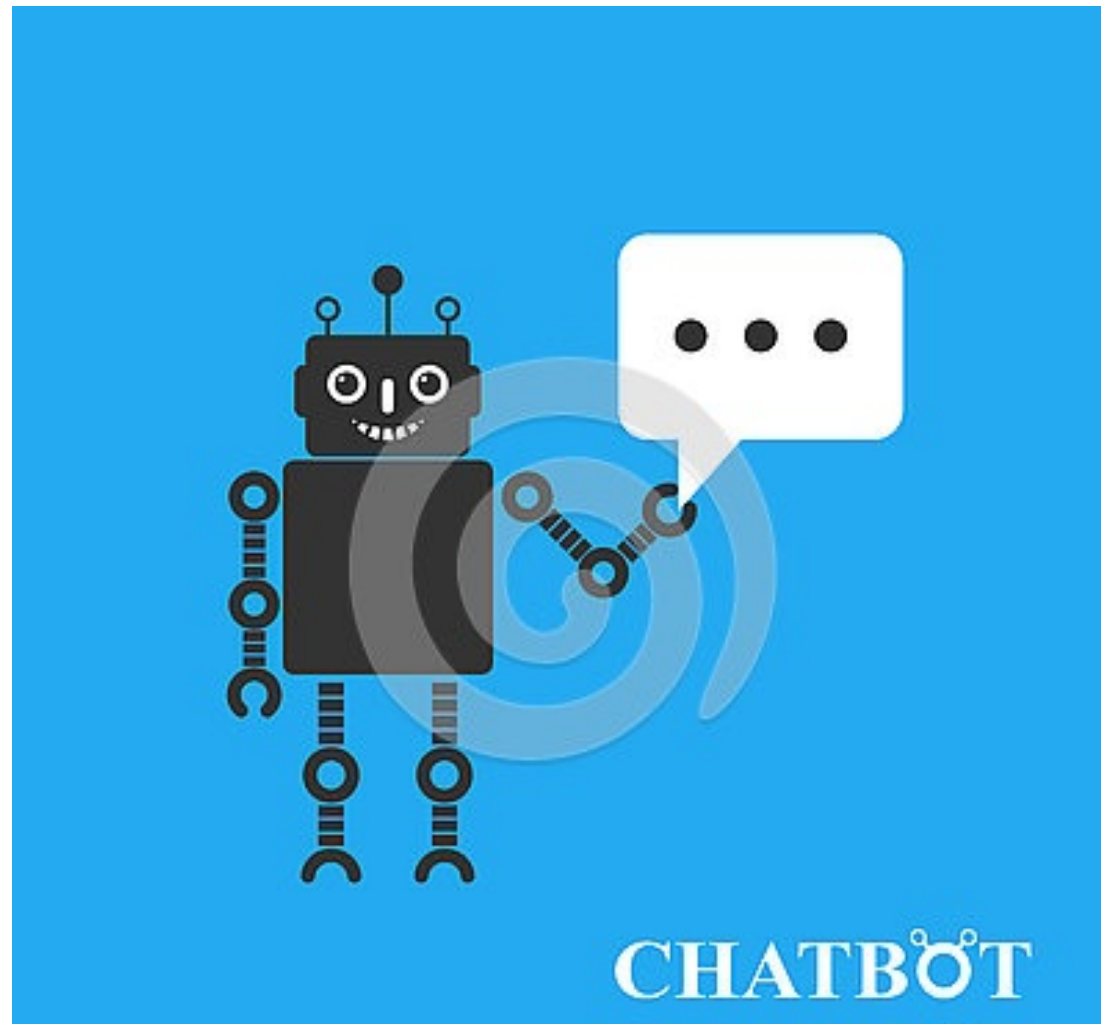
- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng



Cancer?

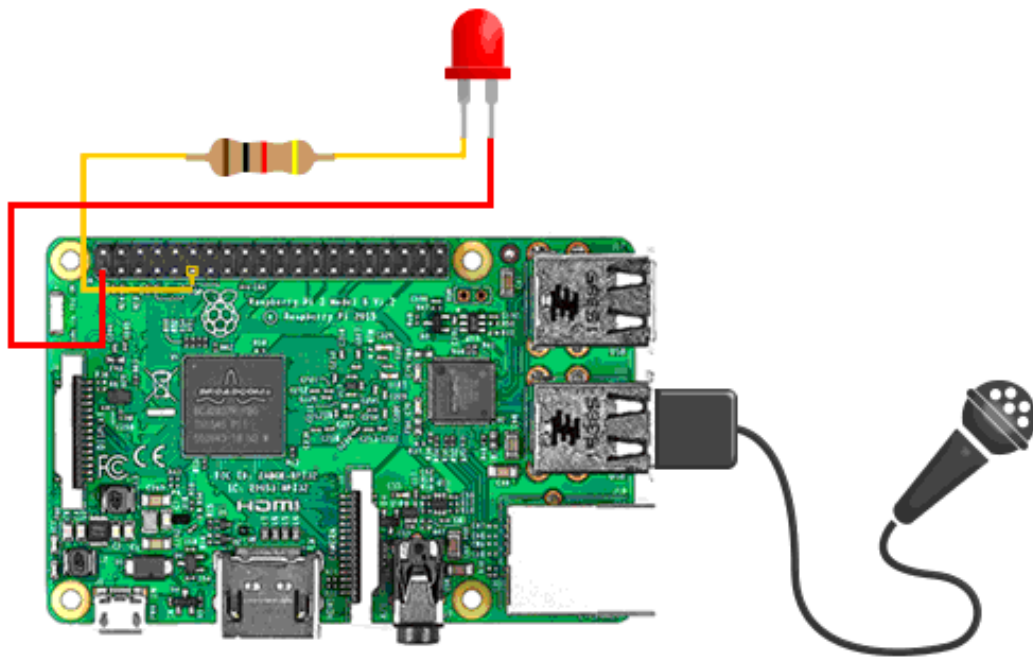
Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng



Ứng dụng

- Tại sao máy học là cần thiết?
- Máy học và các bài toán
- Ứng dụng





Merci !