



TRƯỜNG ĐẠI HỌC  
BÁCH KHOA HÀ NỘI  
HANOI UNIVERSITY  
OF SCIENCE AND TECHNOLOGY

# IT3160

## Nhập môn Trí tuệ nhân tạo

*Artificial Intelligence*

Th.S Ngô Văn Linh

Trường Công nghệ thông tin và Truyền thông

Đại học Bách khoa Hà Nội

ONE LOVE. ONE FUTURE.

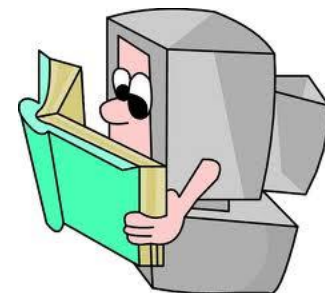


# Nội dung môn học

- Chương 1. Tổng quan
- Chương 2. Tác tử thông minh
- Chương 3. Giải quyết vấn đề
- Chương 4. Tri thức và suy diễn
- **Chương 5. Học máy**
  - **Giới thiệu về học máy**
  - K láng giềng gần
  - Phân lớp Naïve Bayes

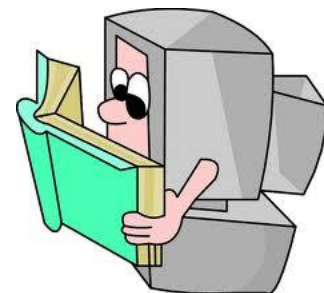
# Giới thiệu về Học máy

- Học máy (ML - Machine Learning) là một lĩnh vực nghiên cứu của Trí tuệ nhân tạo (Artificial Intelligence)
- Câu hỏi trung tâm của ML:
  - “*How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?*”  
[Mitchell, 2006]
- Vài quan điểm về học máy:
  - Một quá trình nhờ đó một hệ thống cải thiện hiệu suất (hiệu quả hoạt động) của nó [Simon, 1983]
  - Việc lập trình các máy tính để tối ưu hóa một tiêu chí hiệu suất dựa trên các dữ liệu hoặc kinh nghiệm trong quá khứ [Alpaydin, 2010]



# Một máy học

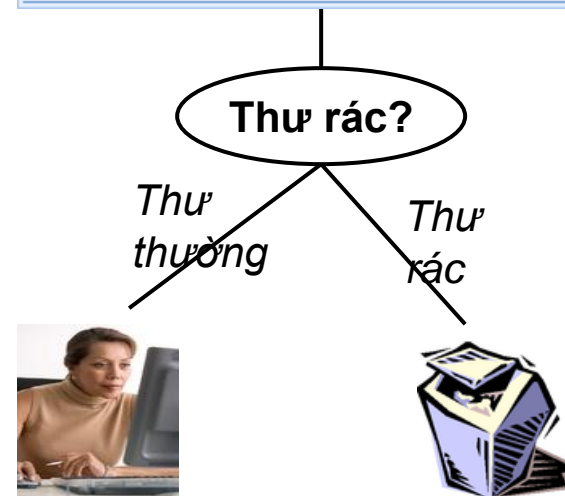
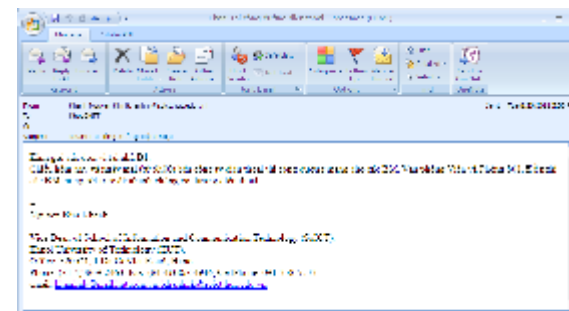
- Ta nói một máy tính *có khả năng học* nếu nó tự cải thiện hiệu suất hoạt động  $P$  cho một công việc  $T$  cụ thể, dựa vào kinh nghiệm  $E$  của nó.
- Như vậy *một bài toán học máy* có thể biểu diễn bằng 1 bộ  $(T, P, E)$ 
  - $T$ : một công việc (nhiệm vụ)
  - $P$ : tiêu chí đánh giá hiệu năng
  - $E$ : kinh nghiệm



# Ví dụ bài toán học máy (1)

## Lọc thư rác (email spam filtering)

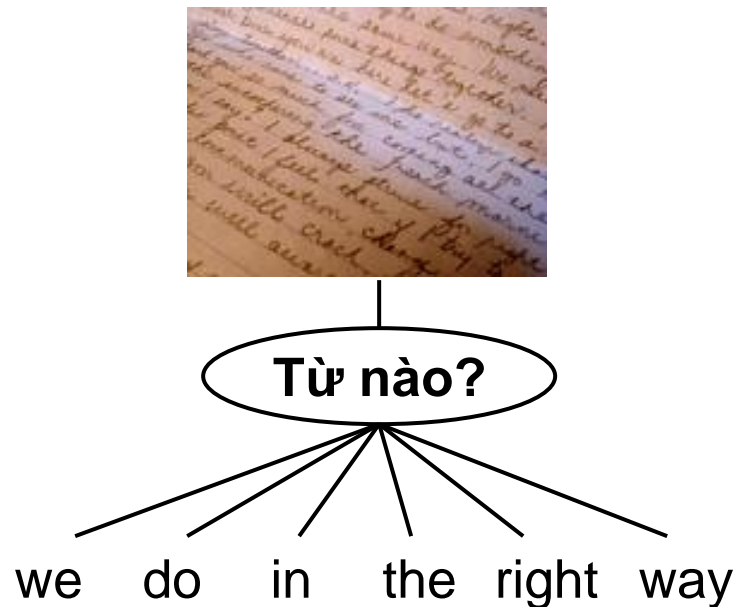
- **T**: Dự đoán (để lọc) những thư điện tử nào là thư rác (spam email)
- **P**: số lượng thư điện tử gửi đến được phân loại chính xác
- **E**: Một tập các thư điện tử (emails) mẫu, mỗi thư điện tử được biểu diễn bằng một tập thuộc tính (vd: tập từ khóa) và nhãn lớp (thư thường/thư rác) tương ứng



# Ví dụ bài toán học máy (2)

## Nhận dạng chữ viết tay

- **T**: Nhận dạng và phân loại các từ trong các ảnh chữ viết
- **P**: Tỷ lệ (%) các từ được nhận dạng và phân loại đúng
- **E**: Một tập các ảnh chữ viết, trong đó mỗi ảnh được gắn với một định danh của một từ



# Ví dụ bài toán học máy (3)

## Gán nhãn ảnh

- **T**: đưa ra một vài mô tả ý nghĩa của 1 bức ảnh
- **P**: ?
- **E**: Một tập các bức ảnh, trong đó mỗi ảnh đã được gán một tập các từ mô tả ý nghĩa của chúng



FISH WATER OCEAN  
TREE CORAL



PEOPLE MARKET PATTERN  
TEXTILE DISPLAY



BIRDS NEST TREE  
BRANCH LEAVES

# Máy học (1)

- Học một ánh xạ (hàm):

$$f : x \mapsto y$$

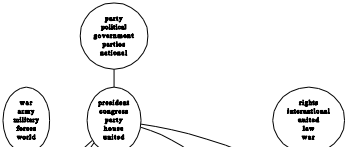
- $x$ : quan sát (dữ liệu), kinh nghiệm
- $y$ : phán đoán, tri thức mới, kinh nghiệm mới, ...
- **Hồi quy** (regression): nếu  $y$  là một số thực
- **Phân loại** (classification): nếu  $y$  thuộc một tập rời rạc (tập nhãn lớp)

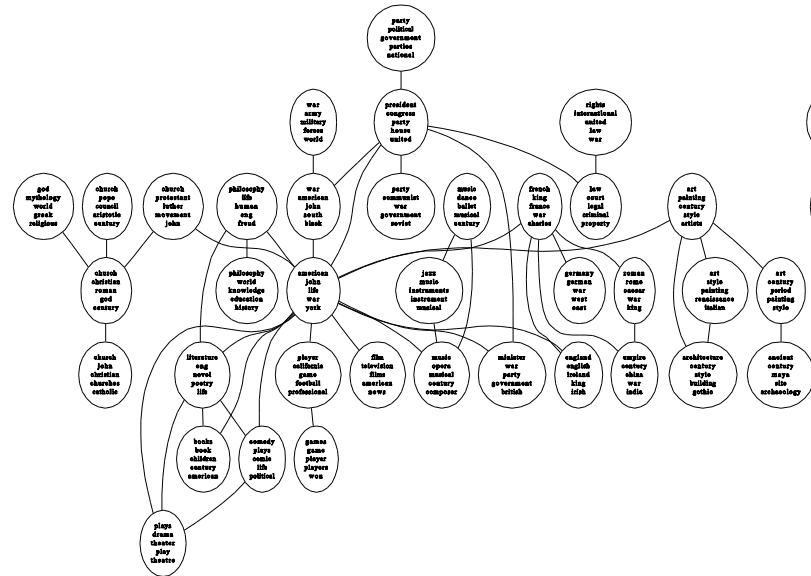


# Máy học (2)

- Học từ đâu?
  - Từ các quan sát trong quá khứ (tập học).  
 $\{X_1, X_2, \dots, X_N\}; \{Y_1, Y_2, \dots, Y_M\}$
- Sau khi đã học:
  - Thu được một mô hình, kinh nghiệm, tri thức mới.
  - Dùng nó để suy diễn (phán đoán) cho quan sát trong tương lai.  
 $Y = f(x)$

# Hai bài toán học cơ bản

- **Học có giám sát (supervised learning):** cần học một hàm  $y = f(x)$  từ tập học  $\{\{x_1, x_2, \dots, x_N\}; \{y_1, y_2, \dots, y_N\}\}$  sao cho  $y_i = f(x_i)$ .
    - *Phân loại* (phân lớp): nếu  $y$  chỉ nhận giá trị từ một tập rời rạc, chẳng hạn {cá, cây, quả, mèo}
    - *Hồi quy*: nếu  $y$  nhận giá trị số thực
  - **Học không giám sát (unsupervised learning):** cần học một hàm  $y = f(x)$  từ tập học cho trước  $\{x_1, x_2, \dots, x_N\}$ .
    - $Y$  có thể là các cụm dữ liệu.
    - $Y$  có thể là các cấu trúc ẩn.
- 



## Học có giám sát: ví dụ

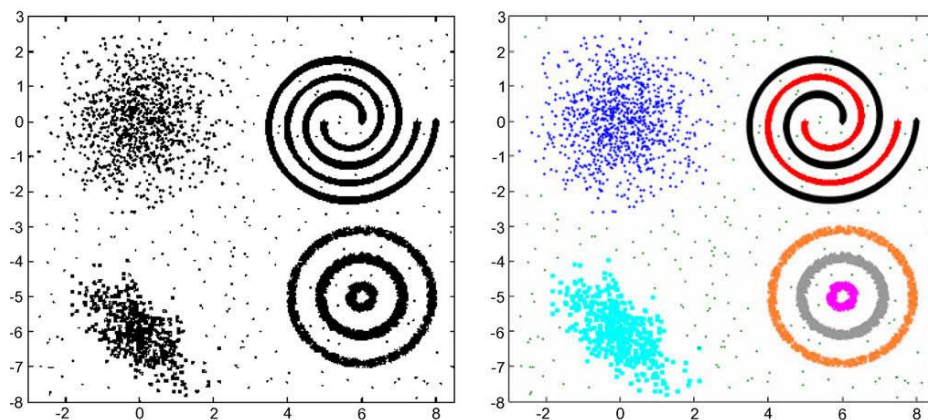
- Lọc thư rác
- Phân loại trang web
- Dự đoán rủi ro tài chính
- Dự đoán biến động chỉ số chứng khoán
- Phát hiện tấn công mạng



# Học không giám sát: ví dụ (1)

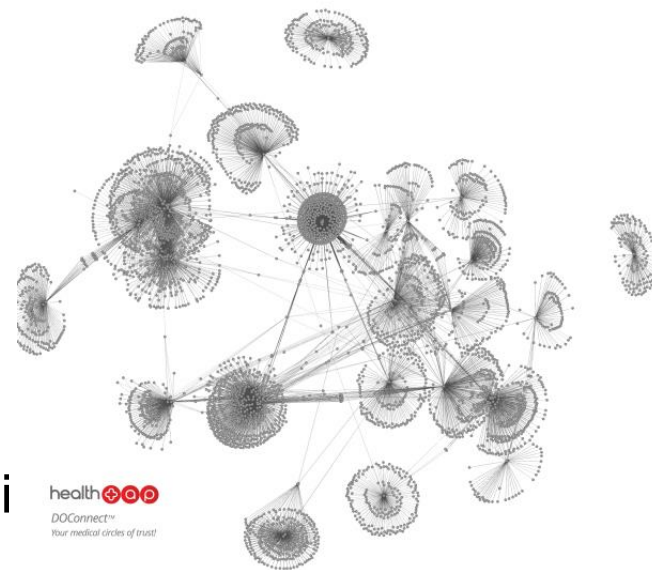
## ■ Phân cụm (clustering)

- Phát hiện các cụm dữ liệu, cụm tính chất,...



## ■ Community detection

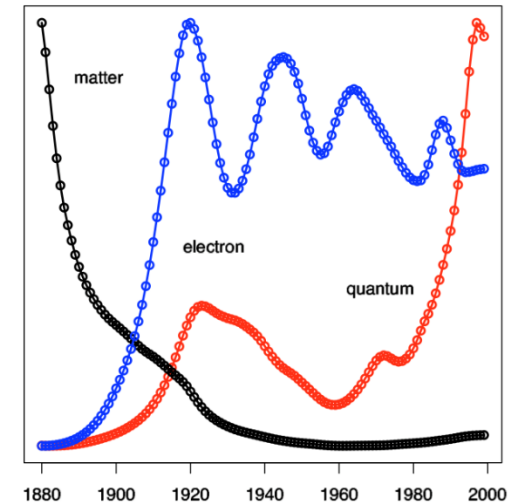
- Phát hiện các cộng đồng trong mạng xã hội



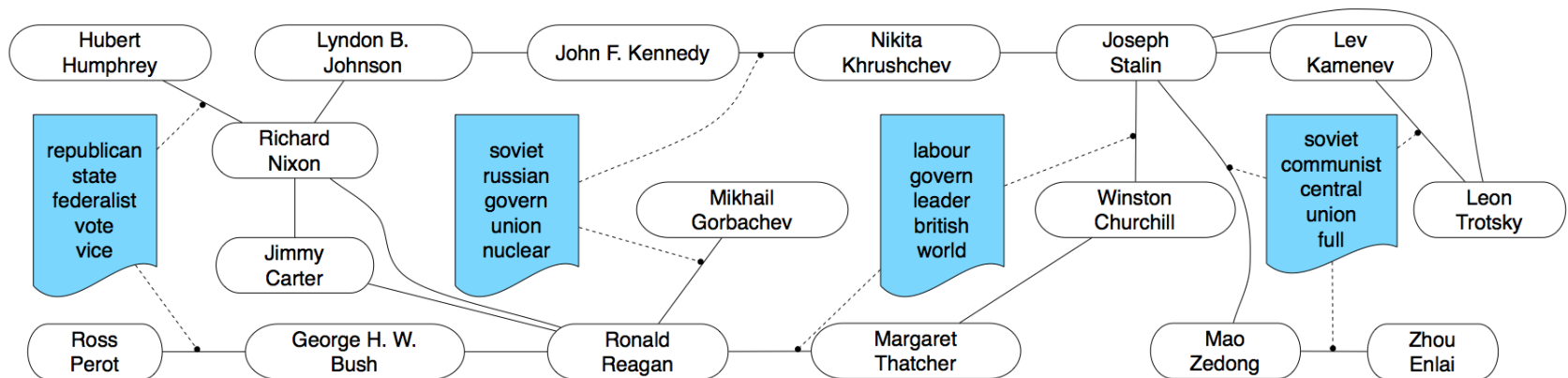
# Học không giám sát: ví dụ (2)

## ■ Trends detection

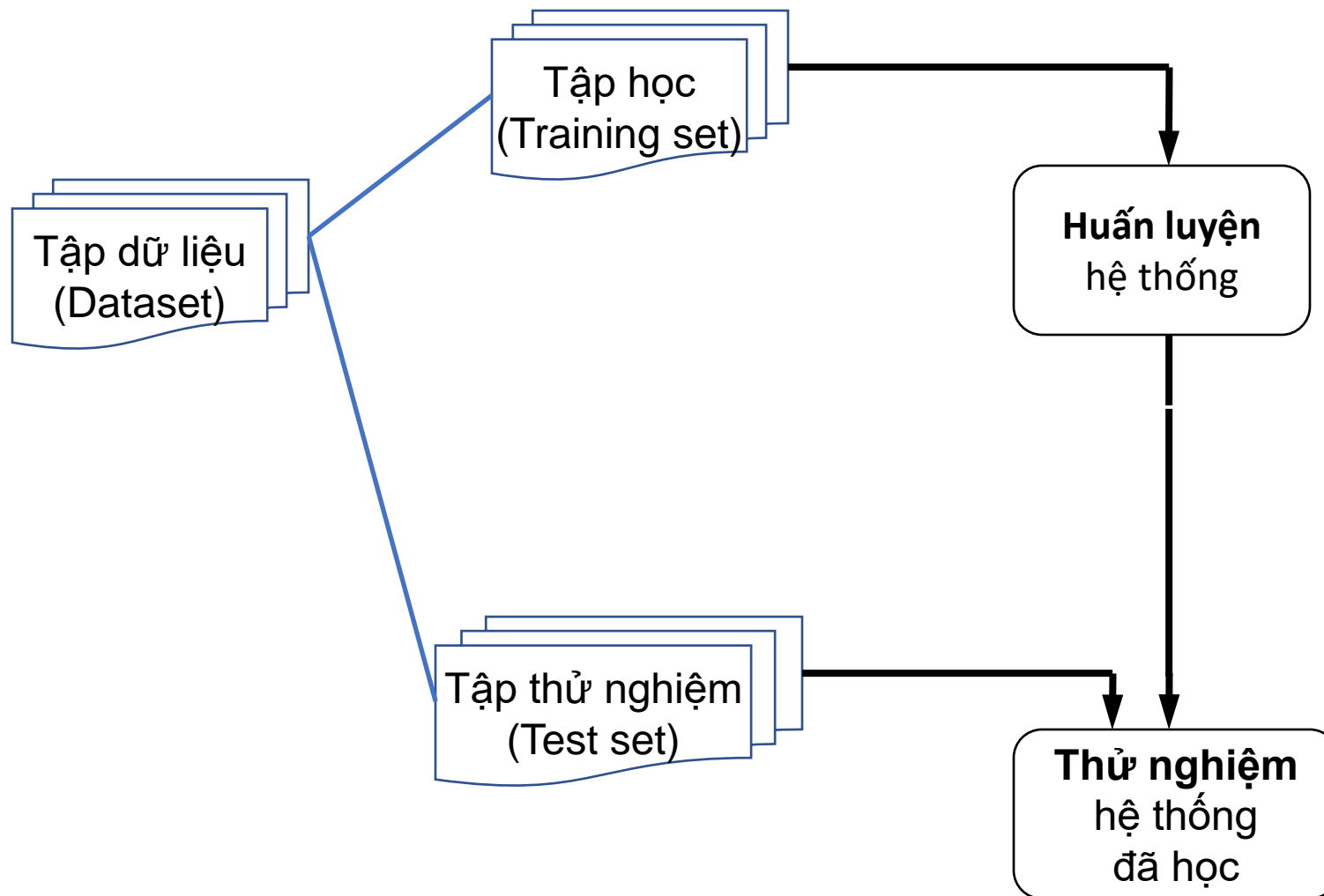
- Phát hiện xu hướng, thị yếu,...



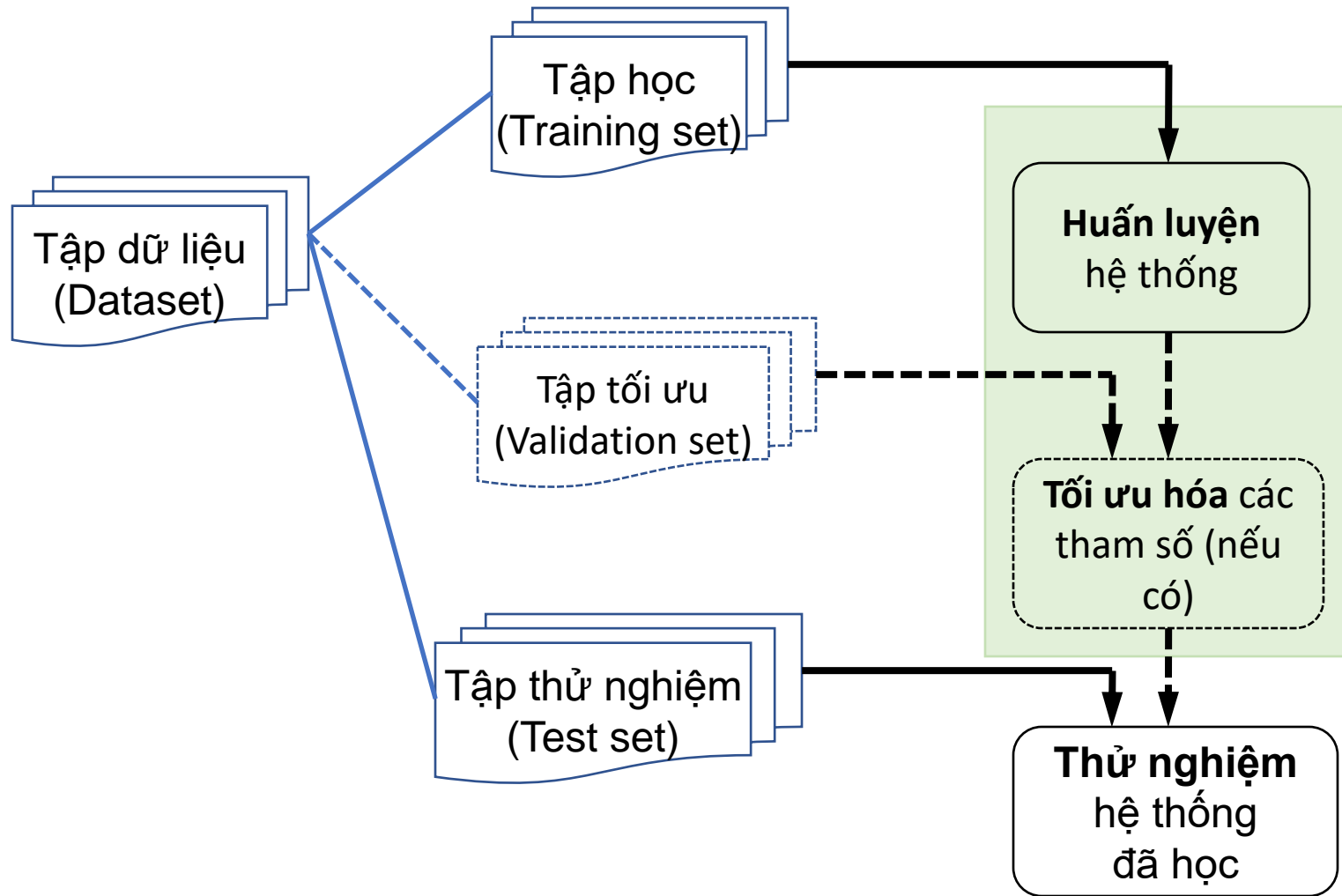
## ■ Entity-interaction analysis




# Quá trình học máy: cơ bản



# Quá trình học máy: toàn diện



# Thiết kế một hệ thống học (1)

- Lựa chọn các ví dụ học (training/learning examples)
  - Các thông tin hướng dẫn quá trình học (training feedback) được chứa ngay trong các ví dụ học, hay là được cung cấp gián tiếp (vd: từ môi trường hoạt động)
  - Các ví dụ học theo kiểu **có giám sát** (supervised) hay không có giám sát (unsupervised)
  - Các ví dụ học nên tương thích với (đại diện cho) các ví dụ sẽ được làm việc bởi hệ thống trong tương lai (future test examples)
- Xác định hàm mục tiêu (giả thiết, khái niệm) cần học
  - $F: X \rightarrow \{0,1\}$
  - $F: X \rightarrow$  Một tập các nhãn lớp 
  - $F: X \rightarrow \mathbb{R}^+$  (miền các giá trị số thực dương)
  - ...



# Thiết kế một hệ thống học (2)

- Lựa chọn cách biểu diễn cho hàm mục tiêu cần học
  - Hàm đa thức (a polynomial function)
  - Một tập các luật (a set of rules)
  - Một cây quyết định (a decision tree)
  - Một mạng nơ-ron nhân tạo (an artificial neural network)
  - ...
- Lựa chọn một giải thuật học máy có thể học (xấp xỉ) được hàm mục tiêu
  - Phương pháp học hồi quy (Regression-based)
  - Phương pháp học quy nạp luật (Rule induction)
  - Phương pháp học cây quyết định (ID3 hoặc C4.5)
  - Phương pháp học lan truyền ngược (Back-propagation)
  - ...



# Các vấn đề trong Học máy (1)

- Giải thuật học máy (Learning algorithm)
  - Những giải thuật học máy nào có thể học (xấp xỉ) một hàm mục tiêu cần học?
  - Với những điều kiện nào, một giải thuật học máy đã chọn sẽ hội tụ (tiệm cận) hàm mục tiêu cần học?
  - Đối với một lĩnh vực bài toán cụ thể và đối với một cách biểu diễn các ví dụ (đối tượng) cụ thể, giải thuật học máy nào thực hiện tốt nhất?

# Các vấn đề trong Học máy (2)

- Các ví dụ học (Training examples)
  - Bao nhiêu ví dụ học là đủ?
  - Kích thước của tập học (tập huấn luyện) ảnh hưởng thế nào đối với độ chính xác của hàm mục tiêu học được?
  - Các ví dụ lỗi (nhiều) và/hoặc các ví dụ thiếu giá trị thuộc tính (missing-value) ảnh hưởng thế nào đối với độ chính xác?



# Các vấn đề trong Học máy (3)

- Quá trình học (Learning process)
  - Chiến lược tối ưu cho việc lựa chọn thứ tự sử dụng (khai thác) các ví dụ học?
  - Các chiến lược lựa chọn này làm thay đổi mức độ phức tạp của bài toán học máy như thế nào?
  - Các tri thức cụ thể của bài toán (ngoài các ví dụ học) có thể đóng góp thế nào đối với quá trình học?

# Các vấn đề trong Học máy (4)

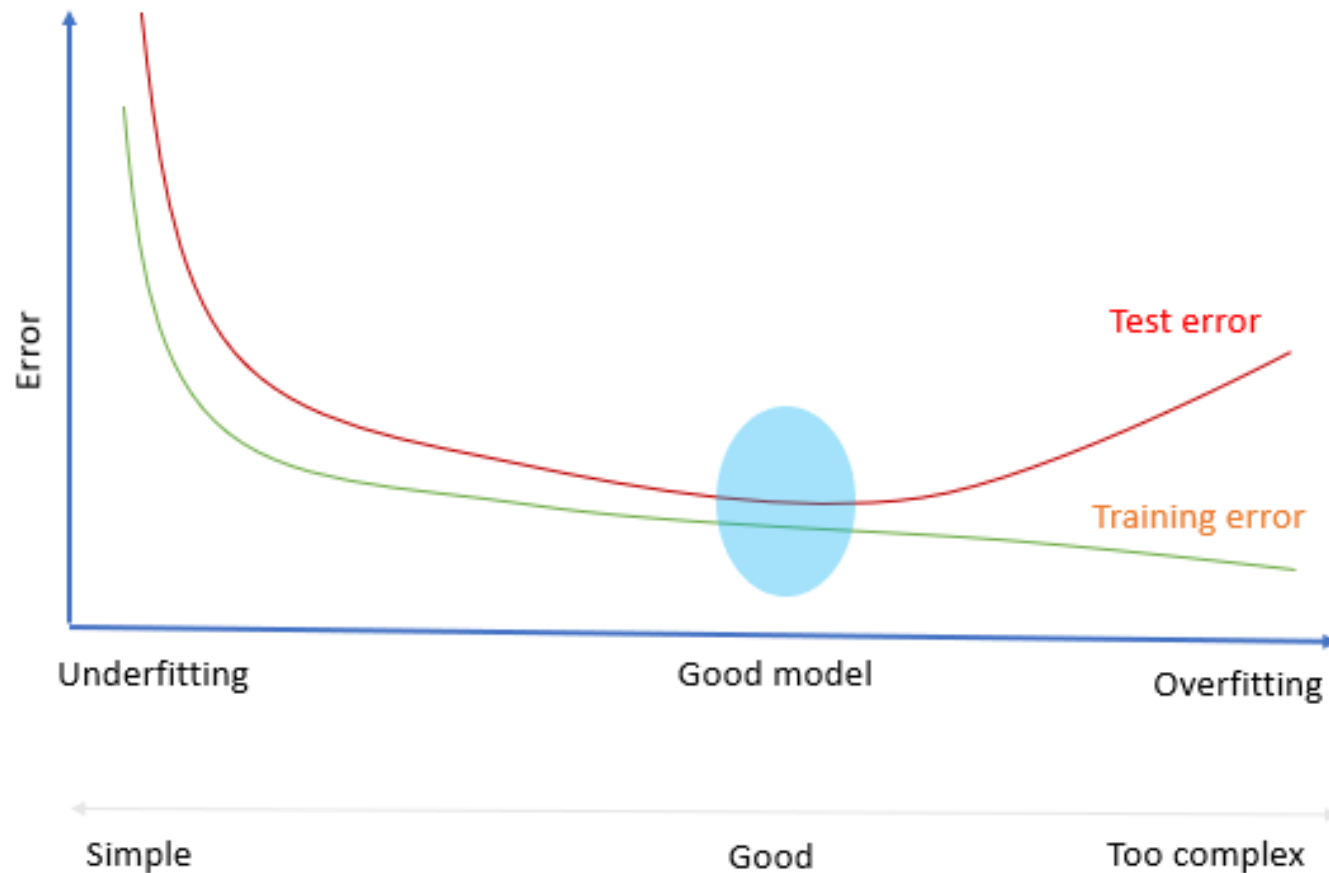
- Khả năng/giới hạn học (Learnability)
  - Hàm mục tiêu nào mà hệ thống cần học?
    - Biểu diễn hàm mục tiêu: Khả năng biểu diễn (vd: hàm tuyến tính / hàm phi tuyến) vs. Độ phức tạp của giải thuật và quá trình học
  - Các giới hạn (trên lý thuyết) đối với khả năng học của các giải thuật học máy?
  - Khả năng khái quát hóa (generalization) của hệ thống?
    - Để tránh vấn đề “over-fitting” (đạt độ chính xác cao trên tập học, nhưng đạt độ chính xác thấp trên tập thử nghiệm)
  - Khả năng hệ thống tự động thay đổi (thích nghi) biểu diễn (cấu trúc) bên trong của nó?
    - Để cải thiện khả năng (của hệ thống đối với việc) biểu diễn và học hàm mục tiêu

# Overfitting (quá khớp, quá khít)

- Hàm  $h$  được gọi là *overfitting* nếu tồn tại hàm  $g$  mà:
  - $g$  có thể tồi hơn  $h$  đối với tập huấn luyện,
  - nhưng  $g$  tốt hơn  $h$  đối với dữ liệu tương lai.
- A learning algorithm is said to overfit relative to another one if it is *more accurate in fitting* known data, but *less accurate in predicting* unseen data.
- Vài nguyên nhân gây ra Overfitting:
  - Hàm  $h$  quá phức tạp
  - Lỗi (nhiều) trong tập huấn luyện (do quá trình thu thập/xây dựng tập dữ liệu)
  - Số lượng các ví dụ học quá nhỏ, không đại diện cho toàn bộ tập (phân bố của các ví dụ của bài toán học)

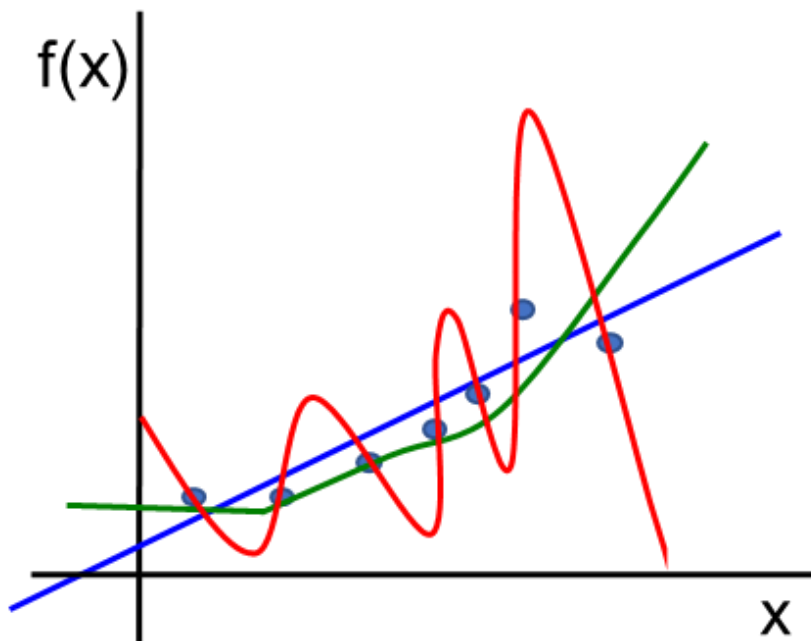


# Vấn đề overfitting: minh họa



# Overfitting

- Trong số rất nhiều hàm thì hàm nào có khả năng tổng quát cao nhất khi học từ tập dữ liệu cho trước?
  - *Tổng quát hoá là mục tiêu chính của học máy.*
  - Tức là, khả năng phán đoán tốt với dữ liệu tương lai.





# Tài liệu tham khảo

- E. Alpaydin. *Introduction to Machine Learning*. The MIT Press, 2010.
- T. M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
- T. M. Mitchell. *The discipline of machine learning*. CMU technical report, 2006.
- H. A. Simon. *Why Should Machines Learn?* In R. S. Michalski, J. Carbonell, and T. M. Mitchell (Eds.): *Machine learning: An artificial intelligence approach*, chapter 2, pp. 25-38. Morgan Kaufmann, 1983.
- A. Kontorovich and Weiss. *A Bayes consistent 1-NN classifier*. Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS). JMLR: W&CP volume 38, 2015.
- A. Guyader, N. Hengartner. *On the Mutual Nearest Neighbors Estimate in Regression*. Journal of Machine Learning Research 14 (2013) 2361-2376.
- L. Gottlieb, A. Kontorovich, and P. Nisnevitch. *Near-optimal sample compression for nearest neighbors*. Advances in Neural Information Processing Systems, 2014.

