HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

KHÓA HỌC

HỌC MÁY CHO AN TOÀN THÔNG TIN

PHẦN 1: CƠ BẢN VỀ HỌC MÁY VÀ HỌC SÂU

Giảng viên: TS. Nguyễn Ngọc Điệp

E-mail: diepnguyenngoc@ptit.edu.vn

Đơn vị: Khoa An toàn thông tin, Học viện Công nghệ BCVT

Nội dung

- Giới thiệu chung về Học máy (ML) và Học sâu (DL)
  - Các khái niệm cơ bản
    - AI, ML, DL
    - Giới thiệu về mô hình học máy
  - Phân loại các mô hình học máy
    - Supervised learning, Unsupervised learning, Semi-supervised learning
    - Reinforcement learning, Transfer learning, Ensemble learning
  - Ứng dụng của học máy

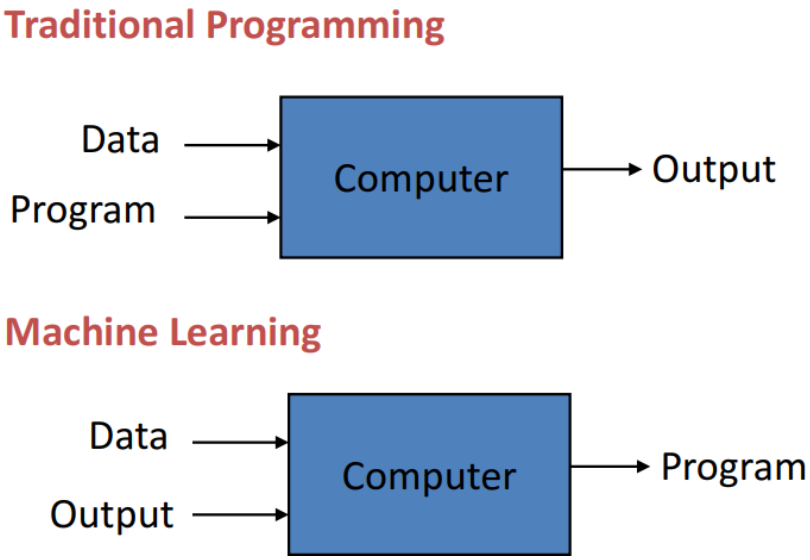
1.1. Các khái niệm cơ bản

- Học máy là gì?
  - Học:
    - …thu thập kiến thức hoặc kỹ năng…
    - Tom Mitchell (1997)
      - Học máy là một lĩnh vực nghiên cứu về việc phát triển các thuật toán máy tính có khả năng tự động cải thiện hiệu suất khi thực hiện một nhiệm vụ thông qua kinh nghiệm.
      - *"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."*
  - Học máy:
    - Giải quyết vấn đề từ kinh nghiệm
    - …được thực hiện bởi chương trình máy tính có khả năng:
      - Thực hiện công việc T tốt hơn
      - Theo tiêu chí P
      - Nhờ sử dụng dữ liệu mẫu hoặc kinh nghiệm E

1.1. Các khái niệm cơ bản

❖ Học máy là gì?

- "Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed." -Arthur Samuel (1959)
- "Học máy: Lĩnh vực nghiên cứu giúp máy tính có khả năng học hỏi mà không cần được lập trình một cách rõ ràng."



Tham khảo slide: Pedro Domingos

Ví dụ

- Học đánh cờ
  - T: đánh cờ
  - P: số ván thắng
  - E: kinh nghiệm tự chơi
- Học nhận dạng chữ
  - T: nhận dạng chữ cái từ ảnh
  - P: phần trăm chữ nhận dạng đúng
  - E: ảnh số của chữ và chữ tương ứng
- Dịch máy
  - T: dịch một câu tiếng Anh sang tiếng Việt
  - P: độ đo dịch máy (ví dụ số câu đúng, số mệnh đề đúng,…)
  - E: cặp câu tiếng Anh và tiếng Việt tương ứng

Vấn đề cần quan tâm (1/2)

- Kinh nghiệm cụ thể như thế nào?
  - Kinh nghiệm trực tiếp và gián tiếp
    - Trực tiếp: trạng thái cụ thể + nước đi đúng tương ứng
    - Gián tiếp: toàn bộ ván cờ và kết quả
  - Có giám sát (hướng dẫn) và không giám sát
    - Có giám sát
    - Không giám sát
    - Bán giám sát
- Cần phải học cái gì? Biểu diễn kiến thức học được thế nào?
  - Tri thức cần học được biểu diễn như một hàm đích, cần lựa chọn hàm đích cụ thể
  - Ví dụ đánh cờ:
    - Chọn_nước_đi: $trạng\ thái \rightarrow nước\ đi$
    - Điểm_số: $trạng\ thái \rightarrow điểm\ số$

Vấn đề cần quan tâm (2/2)

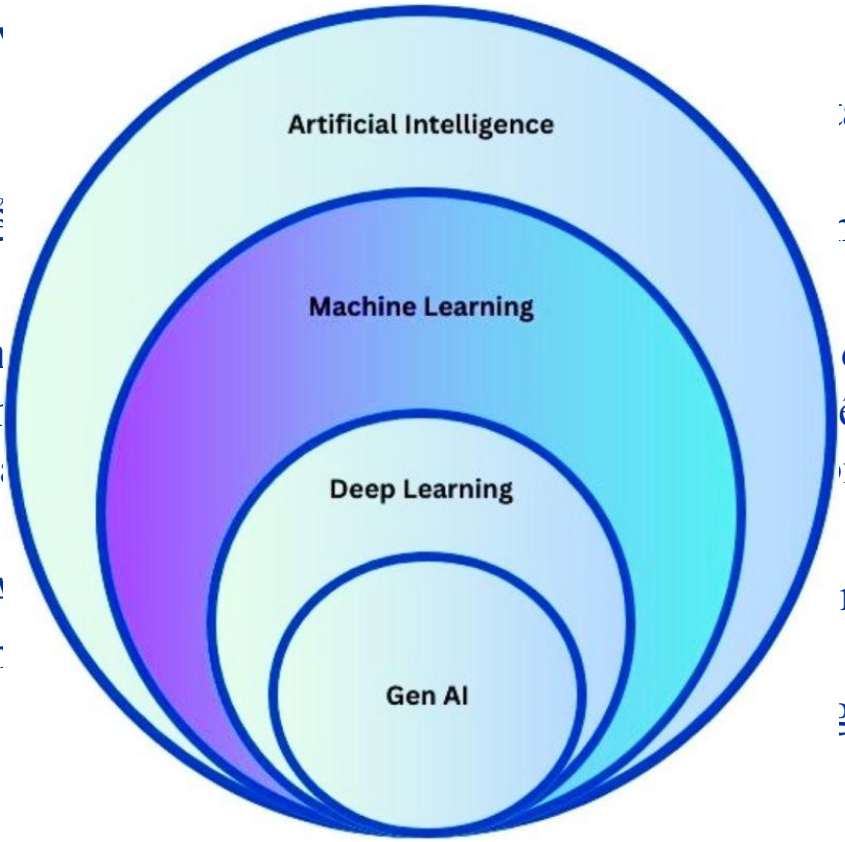
- Sử dụng thuật toán gì để học?
  - Sử dụng hàm
    - VD: $điểm\_số = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + \cdots$
  - Sử dụng các luật
  - Sử dụng mạng nơ ron
  - Sử dụng cây quyết định
  - Sử dụng các mô hình xác suất
  - …

Các khái niệm cơ bản

- Học máy, học sâu và trí tuệ nhân tạo?
  - Học máy (Machine Learning) là một phương pháp trong AI tập trung vào việc phát triển các thuật toán học hỏi từ dữ liệu mà không cần được lập trình cụ thể. Các thuật toán học máy có thể tự động cải thiện hiệu suất khi thực hiện một nhiệm vụ thông qua kinh nghiệm.
  - Học sâu (Deep Learning) là một nhánh con của học máy sử dụng các mạng nơ-ron nhân tạo (artificial neural networks - ANN) để học hỏi từ dữ liệu. ANN được lấy cảm hứng từ cấu trúc của não bộ con người, và có khả năng học hỏi các mô hình phức tạp từ dữ liệu.
  - Trí tuệ nhân tạo (AI) là một lĩnh vực khoa học máy tính rộng lớn bao gồm nhiều ngành nghiên cứu nhằm tạo ra các hệ thống máy móc có khả năng mô phỏng trí tuệ của con người. Học máy và học sâu là hai nhánh con quan trọng của AI.

Các khái niệm cơ bản

- Học máy, học sâu
  - Học máy (Machine ập trung vào việc phát triển các thuật toán trình cụ thể. Các thuật toán học máy có thể một nhiệm vụ thông qua kinh nghiệm.
  - Học sâu (Deep Lea dụng các mạng nơ-ron nhân tạo (artificial n ệu. ANN được lấy cảm hứng từ cấu trúc củ i các mô hình phức tạp từ dữ liệu.
  - Trí tuệ nhân tạo (A ng lớn bao gồm nhiều ngành nghiên cứu n năng mô phỏng trí tuệ của con người. Học g của AI.



Tham khảo từ gupta.sahil.201191/Medium

Các khái niệm cơ bản

❖ Học máy, học sâu và trí tuệ nhân tạo?

Đặc điểm	Trí tuệ nhân tạo (AI)	Học máy (ML)	Học sâu (DL)
Khái niệm	Lĩnh vực rộng lớn bao gồm nhiều ngành nghiên cứu	Nhánh con của AI tập trung vào việc học hỏi từ dữ liệu	Nhánh con của học máy sử dụng mạng nơ-ron nhân tạo
Mục tiêu	Tạo ra các hệ thống máy móc có khả năng mô phỏng trí tuệ của con người	Phát triển các thuật toán học từ dữ liệu	Học các mô hình phức tạp từ dữ liệu
Phương pháp	Sử dụng nhiều kỹ thuật khác nhau, bao gồm học máy, học sâu, xử lý ngôn ngữ tự nhiên, v.v.	Sử dụng các thuật toán học máy để học hỏi từ dữ liệu	Sử dụng mạng nơ-ron nhân tạo để học hỏi từ dữ liệu
Ứng dụng	Rất rộng rãi, bao gồm robot, xe tự lái, nhận dạng giọng nói, v.v.	Phát hiện gian lận, phân tích dữ liệu, dự đoán giá cả, v.v.	Nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên, dịch máy, v.v.

Các ứng dụng tiên tiến

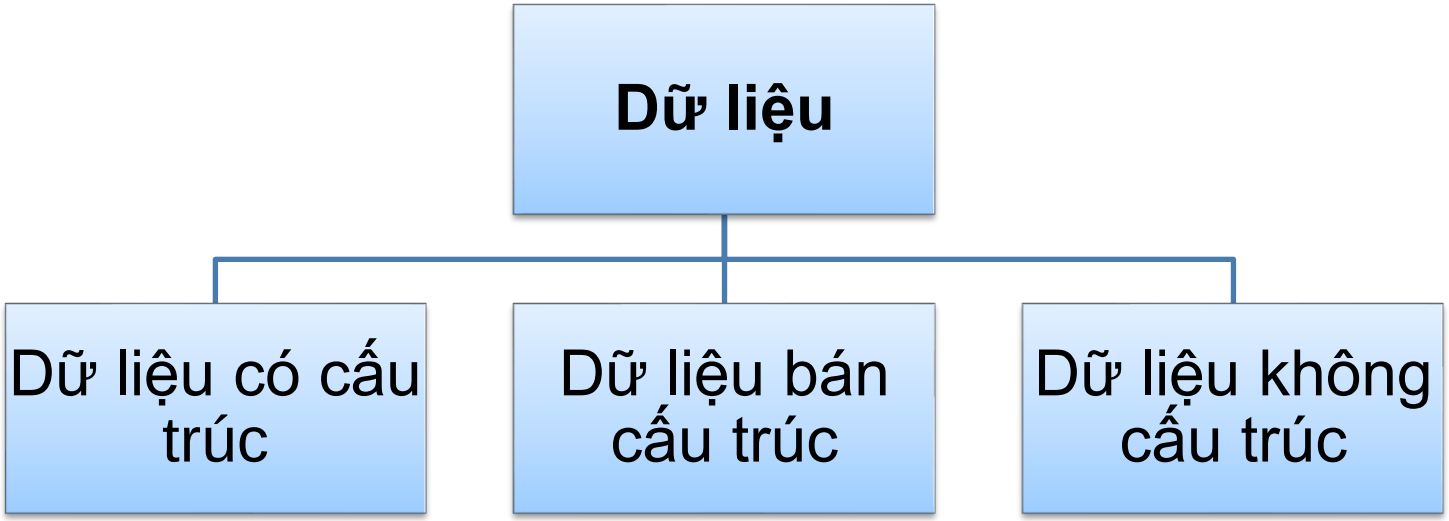
- Xe tự lái:
  - VD: taxi tự lái ở TQ
  - Nhiều cảm biến
- Dịch thuật
- Nhận dạng tiếng nói
- Chữ viết/tiếng nói ➔ video, hình ảnh
- Trợ lý ảo / ChatGPT
- …

Một số khái niệm

- Mẫu, hay ví dụ (samples): là đối tượng cần xử lý (ví dụ phân loại)
  - Ví dụ: khi lọc thư rác thì mỗi thư là một mẫu
- Mẫu thường được mô tả bằng tập thuộc tính hay đặc trưng (features)
  - Ví dụ: trong chuẩn đoán bệnh, thuộc tính là triệu chứng của người bệnh, và các tham số khác như chiều cao, cân nặng, …
- Nhãn phân loại (label): thể hiện loại của đối tượng mà ta cần dự đoán
  - Ví dụ: nhãn phân loại thư rác có thể là “rác” hoặc “bình thường”

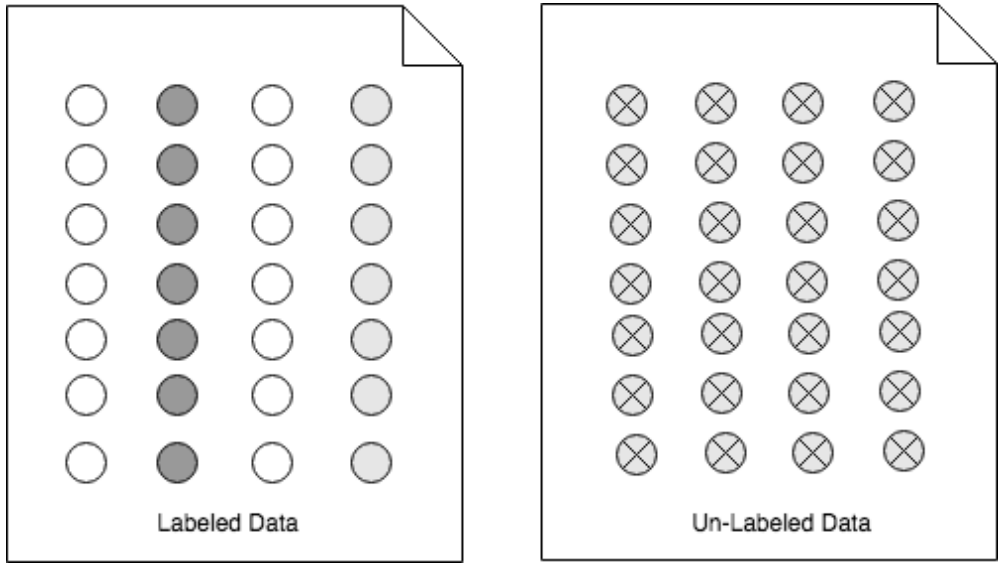
Một số khái niệm

- Dữ liệu trong học máy và phân loại



Một số khái niệm

- Dữ liệu có gán nhãn (labeled data) và không gán nhãn (unlabeled data)



Một số khái niệm

❖ Dữ liệu có gán nhãn và không gán nhãn

- Dữ liệu gán nhãn:
  - dữ liệu được gán nhãn hoặc chú thích để xác định thuộc tính hoặc lớp của nó.
  - sử dụng để huấn luyện các mô hình học máy có giám sát, giúp mô hình học cách thực hiện một nhiệm vụ cụ thể bằng cách cung cấp cho nó các ví dụ đầu vào và đầu ra mong muốn.
  - Ví dụ:
    - Dữ liệu hình ảnh với nhãn "mèo", "chó", "xe hơi", v.v.
    - Dữ liệu văn bản với nhãn "tích cực", "tiêu cực", "trung lập", v.v.
    - Dữ liệu âm thanh với nhãn "nhạc pop", "nhạc rock", "âm thanh môi trường", v.v.

Một số khái niệm

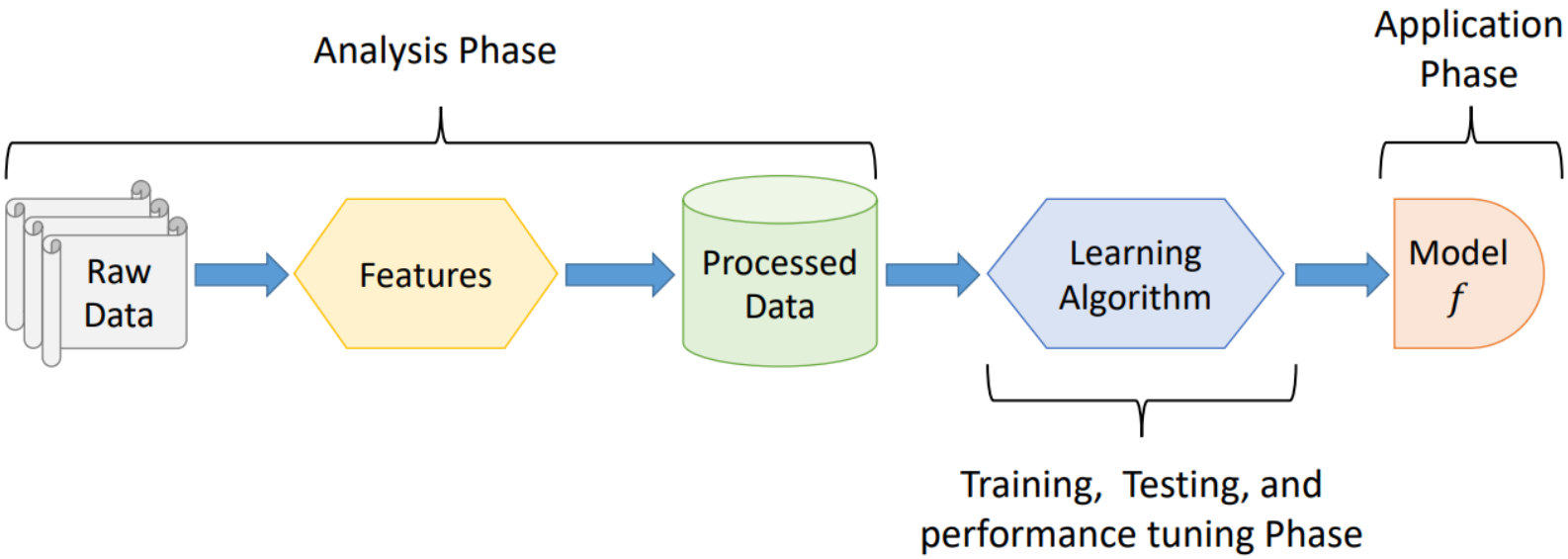
- Dữ liệu có gán nhãn và không gán nhãn
  - Dữ liệu không gán nhãn:
    - dữ liệu không gán nhãn không bao gồm bất kỳ nhãn nào.
    - sử dụng cho các nhiệm vụ học máy không giám sát, mục tiêu là khám phá cấu trúc hoặc thông tin ẩn trong dữ liệu.
    - Ví dụ:
      - Một tập hợp hình ảnh không được gán nhãn
      - Một tập hợp tài liệu văn bản không được gán nhãn
      - Dữ liệu giao dịch không được gán nhãn

Giới thiệu về mô hình học máy

- Mô hình học máy:
  - là một thuật toán được đào tạo dựa trên dữ liệu để học hỏi và thực hiện các nhiệm vụ cụ thể.
  - hoạt động như một "bộ não nhân tạo" có khả năng tự cải thiện thông qua quá trình tiếp xúc với dữ liệu và không cần được lập trình sẵn các quy tắc cụ thể.
- Cách thức hoạt động:
  - Thu thập dữ liệu
  - Lựa chọn thuật toán
  - Đào tạo mô hình
  - Đánh giá và điều chỉnh
  - Sử dụng mô hình
- Lợi ích:
  - Tự động hóa
  - Độ chính xác
  - Khả năng mở rộng
  - Cải tiến liên tục

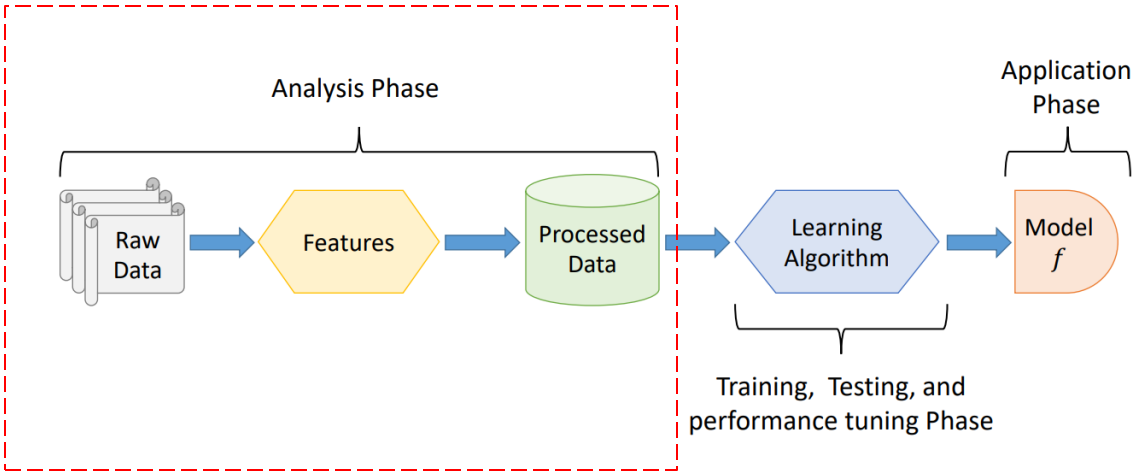
Một số khái niệm

❖ Các pha của học máy



Một số khái niệm

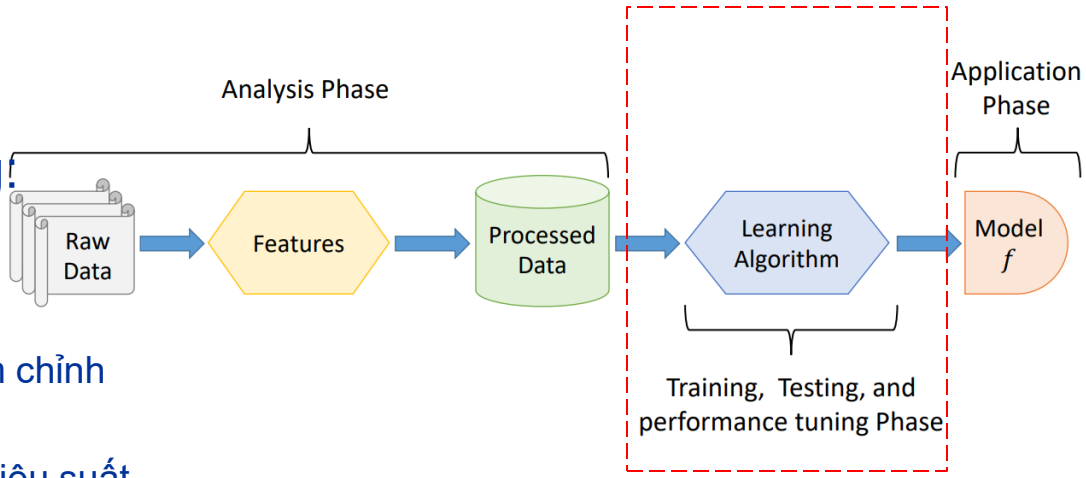
- Các pha của học máy:
  - Pha phân tích dữ liệu
    - Nhằm: Xác định các đặc điểm và mối liên hệ trong dữ liệu để chuẩn bị cho quá trình học máy
    - Gồm:
      - Thu thập dữ liệu
      - Làm sạch dữ liệu
      - Chuyển đổi dữ liệu
      - Khám phá dữ liệu



Một số khái niệm

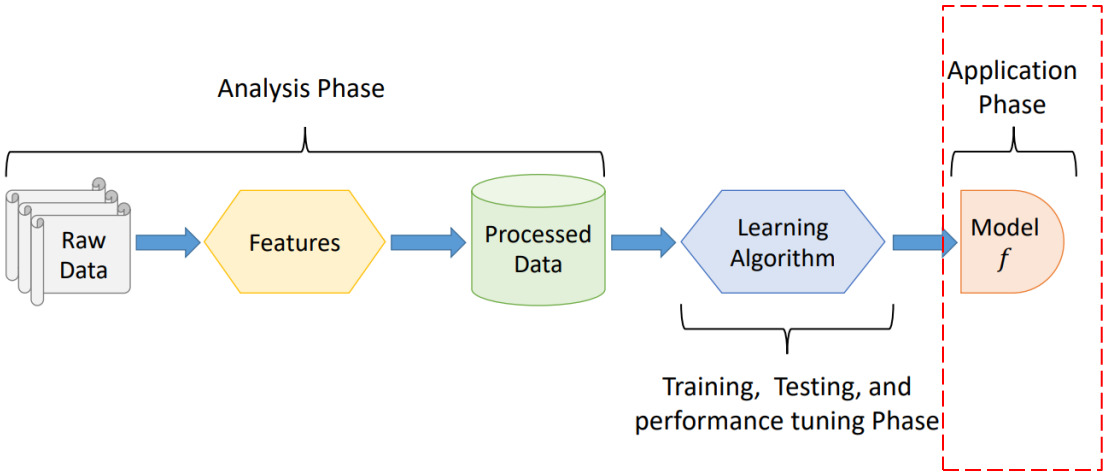
❖ Các pha của học máy

- Pha học máy: gồm Huấn luyện, kiểm thử
  - Nhằm: Huấn luyện mô hình học máy để học hỏi từ dữ liệu và thực hiện các nhiệm vụ cụ thể.
  - Gồm:
    - Lựa chọn thuật toán
    - Chia dữ liệu
    - Huấn luyện mô hình
    - Đánh giá mô hình.
    - Chỉnh sửa mô hình
- Pha tinh chỉnh hiệu năng:
  - Cải thiện hiệu suất
  - Gồm:
    - Lựa chọn và áp dụng pp tinh chỉnh
    - Đánh giá lại mô hình
    - Lặp lại quá trình tinh chỉnh hiệu suất



Một số khái niệm

- Các pha của học máy:
  - Pha ứng dụng:
    - Sử dụng mô hình học máy đã được đào tạo và tinh chỉnh để thực hiện các nhiệm vụ cụ thể trong môi trường thực tế.
    - Gồm:
      - Triển khai mô hình
      - Giám sát mô hình
      - Cập nhật mô hình

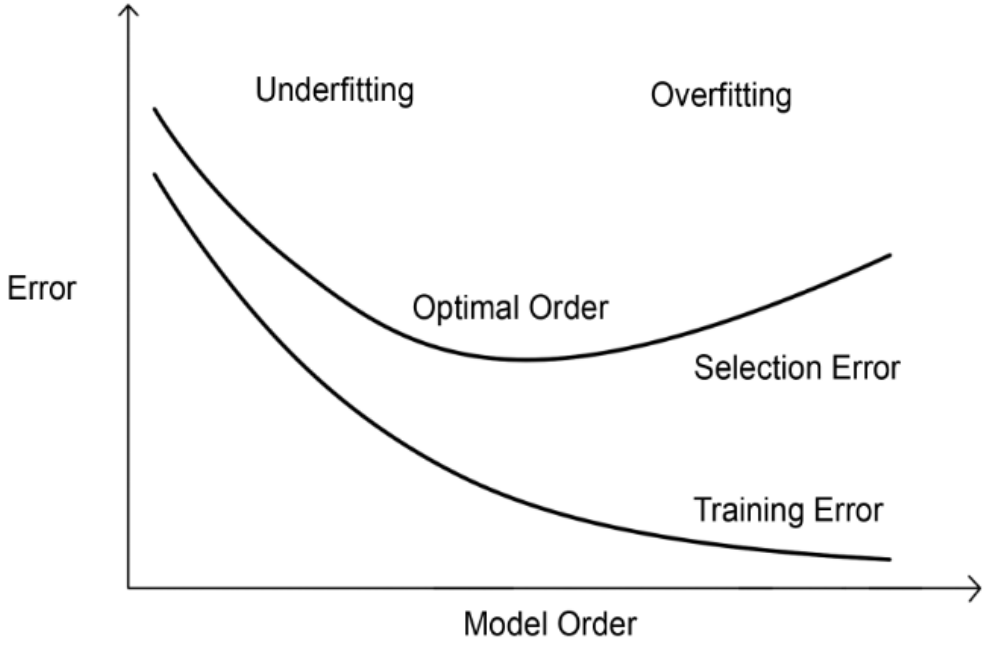


Một số khái niệm

- Sự không nhất quán trong dữ liệu
  - Trong **giai đoạn huấn luyện**, một mô hình học máy có thể hoặc không **tổng quát hóa** hoàn hảo. Điều này là do những sự **không nhất quán trong dữ liệu** cần phải lưu ý.
  - Sự không nhất quán trong dữ liệu là tình trạng dữ liệu sử dụng để huấn luyện mô hình học máy không đồng nhất hoặc có lỗi. Dữ liệu không nhất quán có thể bao gồm các giá trị sai, thiếu sót, hoặc các định dạng khác nhau.
  - Ví dụ:
    - Gán nhãn nhầm email spam và không phải spam
    - Trường diện tích trong dữ liệu dự đoàn giá nhà sử dụng cả đơn vị m^2 và cm^2

Một số khái niệm

- Một số lỗi và vấn đề khi làm việc với mô hình học máy
  - Trong pha huấn luyện
  - Trong pha kiểm tra



Lỗi huấn luyện và Lỗi kiểm tra (1/2)

- Lỗi huấn luyện (Training error)
  - Là lỗi đo được trên tập dữ liệu huấn luyện
  - Thường đo bằng sự sai khác giữa giá trị tính toán của mô hình và giá trị thực của dữ liệu huấn luyện
  - Trong quá trình học ta cố gắng làm giảm tới mức tối thiểu lỗi huấn luyện
- Lỗi kiểm tra (Test error)
  - Là lỗi đo được trên tập dữ liệu kiểm tra
  - Là cái ta thực sự quan tâm

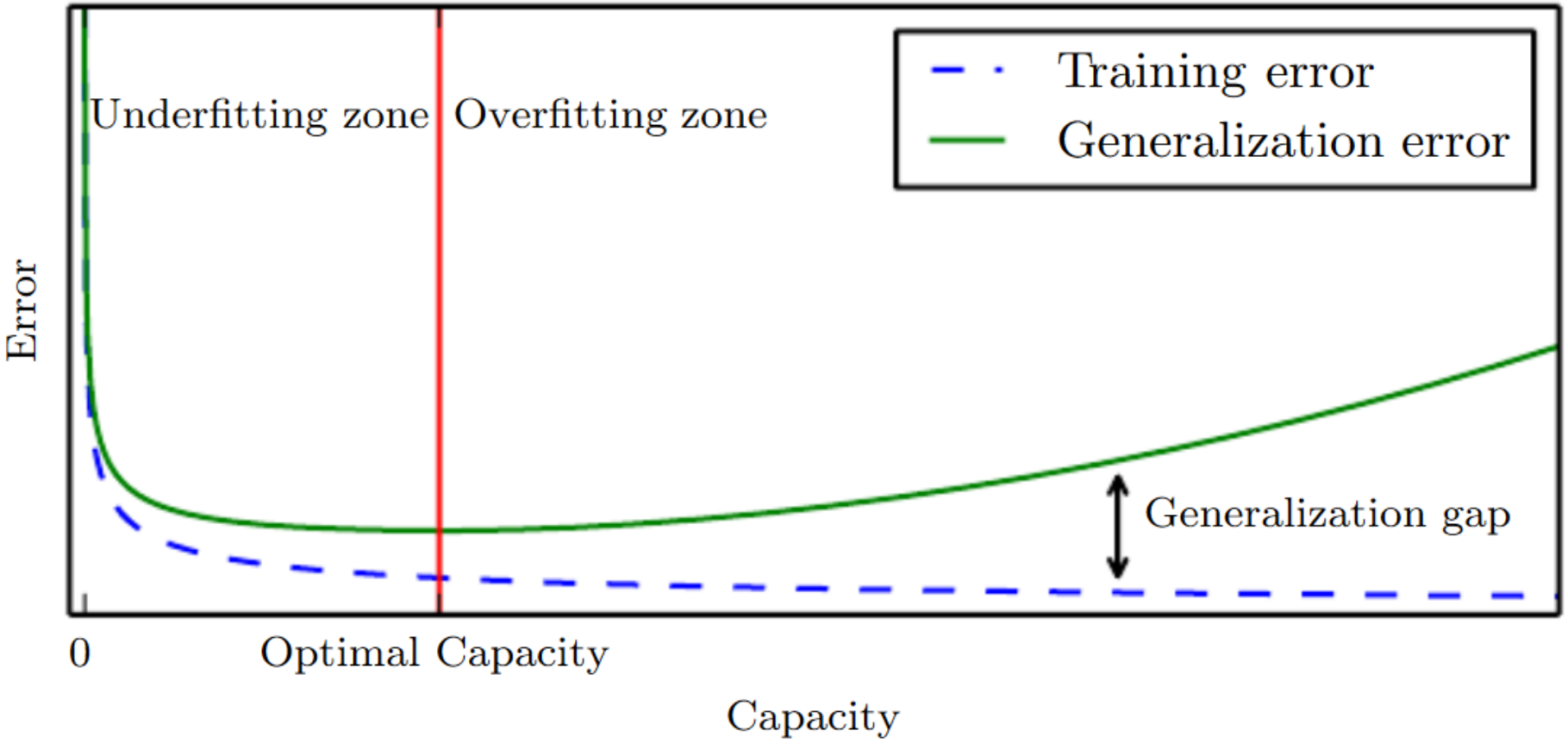
Làm sao ta có thể tác động tới hiệu quả của mô hình trên tập dữ liệu kiểm tra khi ta chỉ quan sát được tập dữ liệu huấn luyện?

Lỗi huấn luyện và Lỗi kiểm tra (2/2)

- i.i.d assumptions (independent, identically distributed)
  - Giả thiết rằng các mẫu dữ liệu (cả ở tập huấn luyện và tập kiểm tra) là độc lập, và các tập dữ liệu huấn luyện và kiểm tra có cùng phân phối
  - Nếu ta cố định các tham số của mô hình thì lỗi huấn luyện và lỗi kiểm tra sẽ bằng nhau
    - Trong quá trình huấn luyện tham số được tối ưu theo lỗi huấn luyện, do đó lỗi kiểm tra thường lớn hơn lỗi huấn luyện

- Hai yếu tố đánh giá độ tốt của một thuật toán học máy
  - Khả năng giảm thiểu lỗi huấn luyện
  - Khả năng giảm thiểu khoảng cách giữa lỗi huấn luyện và lỗi kiểm tra

Underfitting và Overfitting



Underfitting: chưa khớp; Overfitting: quá khớp
Generalization error = test error
Capacity: Khả năng của mô hình

Underfitting và Overfitting

- Chưa khớp (Underfitting):
  - Xảy ra khi mô hình học máy không thể học được đầy đủ các mẫu và mối quan hệ trong dữ liệu huấn luyện.
  - Biểu hiện:
    - Đo lường hiệu suất thấp
    - Dự đoán đơn giản
    - Không linh hoạt
  - Nguyên nhân
    - Dữ liệu huấn luyện quá ít
    - Mô hình quá đơn giản
    - Thiếu tiền xử lý dữ liệu
  - Cách khắc phục:
    - Thêm dữ liệu
    - Làm sạch dữ liệu
    - Sử dụng các kỹ thuật điều chỉnh mô hình
  - Ví dụ: dự đoán giá nhà
    - Nếu bị underfitting, mô hình có thể luôn dự đoán giá nhà là $100.000, bất kể diện tích, vị trí hoặc các yếu tố khác của ngôi nhà

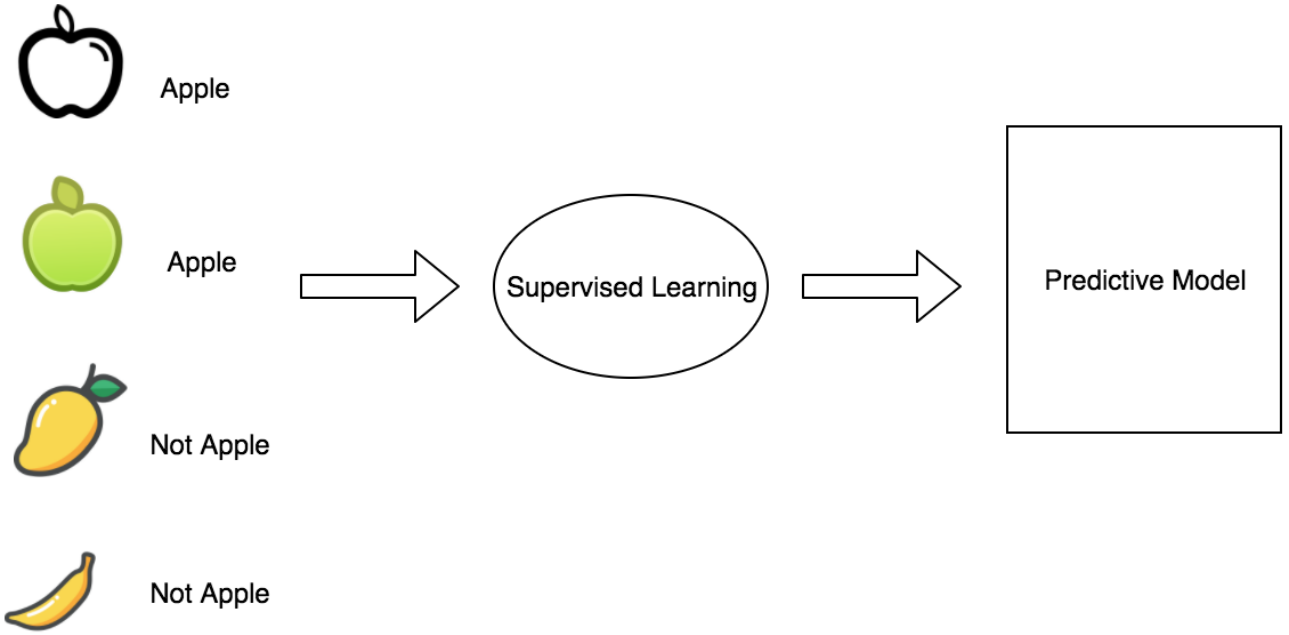
Underfitting và Overfitting

- Quá khớp (Overfitting):
  - Xảy ra khi mô hình học máy không thể học được đầy đủ các mẫu và mối quan hệ trong dữ liệu huấn luyện.
  - Biểu hiện:
    - Hiệu suất cao trên tập dữ liệu huấn luyện
    - Dự đoán phức tạp
    - Không linh hoạt
  - Nguyên nhân
    - Dữ liệu huấn luyện quá ít
    - Mô hình quá phức tạp
    - Thiếu dữ liệu xác thực
  - Cách khắc phục:
    - Thu thập thêm dữ liệu
    - Sử dụng mô hình đơn giản hơn
    - Sử dụng kỹ thuật điều chỉnh mô hình
    - Sử dụng dữ liệu kiểm tra
  - Ví dụ: phân loại email spam
    - Nếu mô hình overfitting, mô hình học cách phân loại email dựa trên các chi tiết cụ thể thay vì dựa trên các yếu tố chung ➔ dự đoán sai

1.2. Phân loại các mô hình học máy

- Học có giám sát (supervised learning)
  - Phân lớp (classification)
  - Hồi quy (regression)
- Học không giám sát (unsupervised learning)
  - Học luật kết hợp (association)
  - Phân cụm (clustering)
- Học bán giám sát (semi-supervised learning)
- Học tăng cường (reinforcement learning)
- Học chuyển giao (transfer learning)
- Học kết hợp (Ensemble learning)

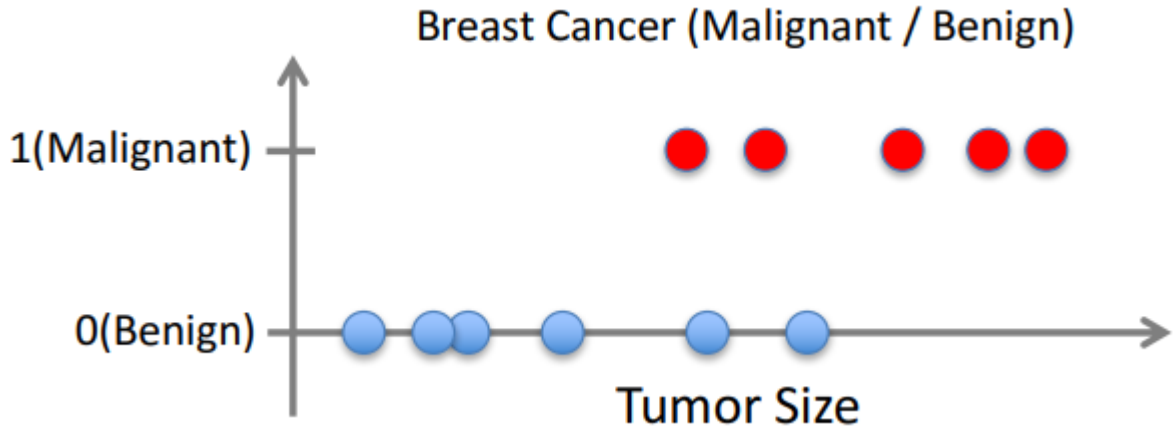
Học có giám sát (supervised learning)



Cho trước: Dữ liệu huấn luyện + đầu ra mong muốn (nhãn)

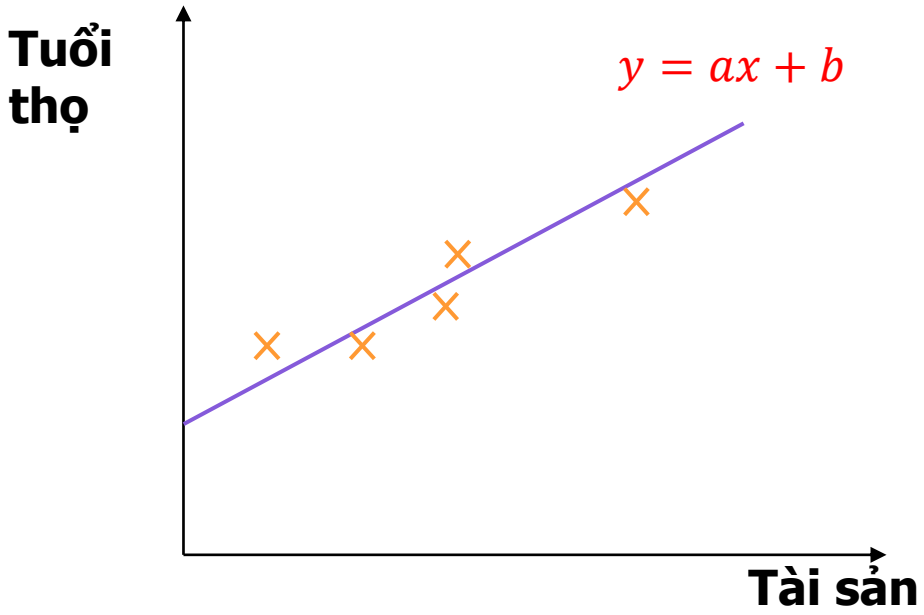
Phân lớp

- Cho $(x_1, y_1), (x_2, y_2),..., (x_n, y_n)$
- Học một hàm f(x) để dự đoán y với x cho trước
  - y là giá trị theo lớp ➔ phân lớp



Hồi quy (regression)

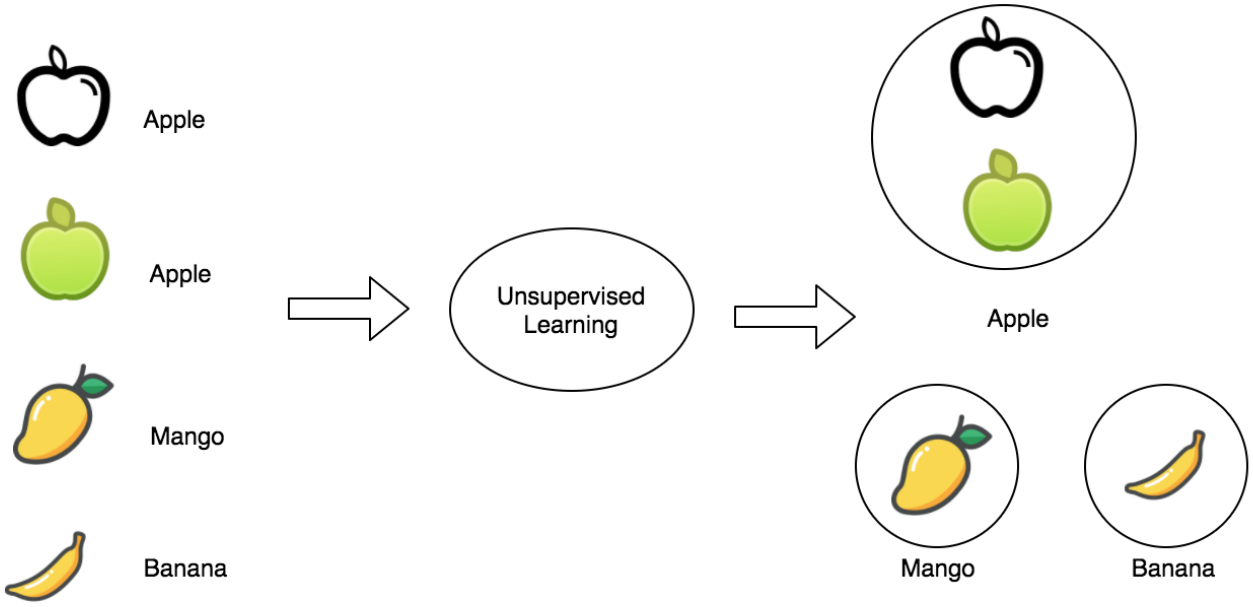
- Cho $(x_1, y_1), (x_2, y_2),..., (x_n, y_n)$
- Học một hàm f(x) để dự đoán y với x cho trước
  - y là giá trị thực ➔ hồi quy



Ứng dụng: dự đoán giá nhà, lái xe,...

Học không giám sát (unsupervised learning)

❖ Có dữ liệu huấn luyện nhưng không có đầu ra mong muốn



Học luật kết hợp

- Ví dụ
  - Phân tích giao dịch, mua bán (hóa đơn mua hàng)
- $P(Y|X)$
  - Xác suất người mua hàng X còn mua hàng Y
- Ví dụ luật kết hợp
  - Người mua bánh mì thường mua bơ
  - Người mua lạc rang thường mua bia

Phân cụm

- Nhóm những trường hợp tương tự với nhau
- Không có giá trị đầu ra
- Ứng dụng
  - Phân cụm khách hàng, phân cụm sinh viên
  - Phân đoạn ảnh
  - Thiết kế vi mạch

Học tăng cường

- Kinh nghiệm không được cho trực tiếp dưới dạng đầu vào / đầu ra
- Hệ thống nhận được một giá trị thưởng (reward) là kết quả cho một chuỗi hành động nào đó
- Thuật toán cần học cách hành động để cực đại hóa giá trị thưởng
- Ví dụ: học đánh cờ
  - Hệ thống không được chỉ cho nước đi nào là hợp lý cho từng tình huống cụ thể
  - Chỉ biết kết quả thắng thua sau một chuỗi nước đi

Phân loại các mô hình học máy

- Học chuyển giao:
  - Sử dụng kiến thức thu được từ việc học một nhiệm vụ (nhiệm vụ nguồn) để cải thiện hiệu suất khi học một nhiệm vụ khác (nhiệm vụ mục tiêu)
    - tận dụng mô hình đã được đào tạo sẵn trên một tập dữ liệu lớn cho một nhiệm vụ cụ thể để áp dụng cho một nhiệm vụ mới có liên quan, thay vì đào tạo mô hình mới hoàn toàn từ đầu
  - Lợi ích:
    - Tiết kiệm thời gian và tài nguyên
    - Cải thiện hiệu suất
    - Khả năng áp dụng cho các tập dữ liệu nhỏ
  - Ứng dụng:
    - Nhận dạng hình ảnh, Xử lý ngôn ngữ tự nhiên, …
    - Ví dụ: bài toán phân loại chó và mèo dựa trên ảnh

Phân loại các mô hình học máy

- Học kết hợp:
  - Kết hợp nhiều mô hình học máy khác nhau để tạo ra một mô hình dự đoán tốt hơn bất kỳ mô hình nào trong số các mô hình thành phần riêng lẻ.
  - Giúp cải thiện độ chính xác, hiệu suất và khả năng tổng quát của mô hình bằng cách kết hợp các điểm mạnh của nhiều mô hình khác nhau.
  - Các phương pháp chính: Bagging, Boosting, Stacking
  - Lợi ích:
    - Cải thiện độ chính xác
    - Giảm nhiễu
    - Cải thiện khả năng tổng quát hóa của mô hình
  - Ứng dụng: nhận dạng ảnh, xử lý ngôn ngữ tự nhiên, …
    - Ví dụ: kết hợp nhiều mô hình để dự đoán giá nhà

1.3. Ứng dụng của học máy

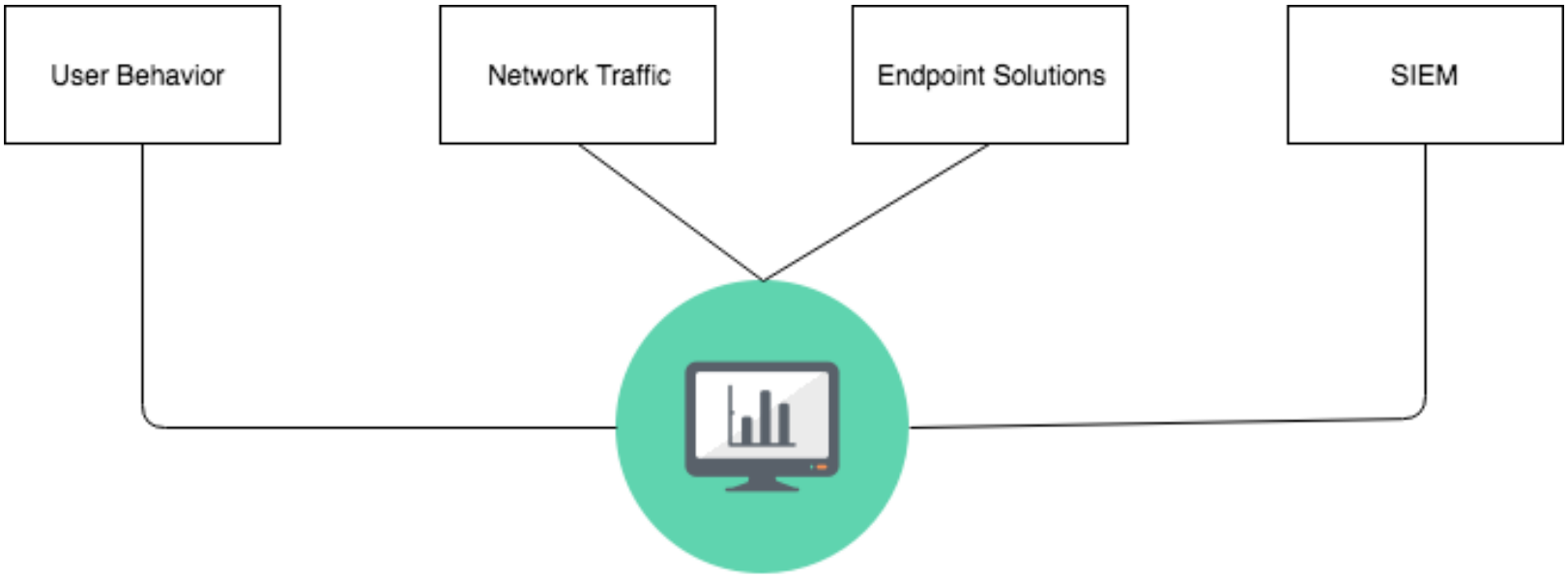
- Những ứng dụng khó lập trình theo cách thông thường do không tồn tại hoặc khó giải thích kinh nghiệm, kỹ năng của con người
  - Nhận dạng chữ viết, âm thanh, hình ảnh
  - Lái xe tự động, thám hiểm sao Hoả
- Chương trình máy tính có khả năng thích nghi: lời giải thay đổi theo thời gian hoặc theo tình huống cụ thể
  - Chương trình trợ giúp cá nhân
  - Định tuyến mạng

- Khai phá (phân tích) dữ liệu
  - Hồ sơ bệnh án → tri thức y học
  - Dữ liệu bán hàng → quy luật kinh doanh

Một số lĩnh vực ứng dụng

Lĩnh vực	Mô tả
Nhận dạng khuôn mặt	Xác định người từ hình ảnh, sử dụng trong hệ thống an ninh.
Phát hiện tin giả	Phân biệt tin giả với tin thật dựa trên nội dung và nguồn gốc.
Phân tích cảm xúc	Xác định mức độ tích cực/tiêu cực của văn bản, đánh giá ý kiến khách hàng.
Hệ tư vấn	Đề xuất sản phẩm/dịch vụ phù hợp dựa trên lịch sử mua sắm và sở thích khách hàng.
Hệ thống phát hiện gian lận	Phát hiện giao dịch bất thường, bảo vệ lợi ích khách hàng.
Dịch máy	Dịch văn bản từ ngôn ngữ này sang ngôn ngữ khác.
Chatbot	Tương tác tự động với khách hàng, hỗ trợ chăm sóc khách hàng và bán hàng.
Trợ lý ảo	Công cụ giao tiếp tự động hỗ trợ con người trong nhiều lĩnh vực: VD ChatGPT, Gemini, …

Một số ứng dụng trong an toàn thông tin



Machine Learning Solutions in Cyber Security

Kết luận

- Giới thiệu chung về Học máy (ML) và Học sâu (DL)
  - Các khái niệm cơ bản
    - AI, ML, DL
  - Giới thiệu về mô hình học máy
- Phân loại các mô hình học máy
  - Supervised learning, Unsupervised learning, Semi-supervised learning
  - Reinforcement learning, Transfer learning, Ensemble learning
- Ứng dụng của học máy
  - Các ứng dụng phổ biến
  - Trong an toàn thông tin