**Tuần 1: TỔNG QUAN VỀ HỌC MÁY**

Học máy (machine learning) là một ngành khoa học nghiên cứu các thuật toán cho phép máy tính có thể học được các khái niệm.

**Phân loại: có 2 loại học máy chính**

* Phương pháp quy nạp: Máy sẽ học được các khái niệm dựa trên dữ liệu đã thu thập được trước đó. Phương pháp này cho phép tận dụng được nguồn dữ liệu rất nhiếu và sẵn có
* Phương pháp suy diễn: Máy sẽ học được các khái niệm dựa vào các nguyên tắc. Phương pháp này cho phép sử dụng các kiến thức chuyên ngành để hỗ trợ máy tính.

**Các dạng học máy :**

* Học có giám sát: Máy tính được cung cấp dữ liệu đầu vào và đầu ra mong muốn. Sau khi học xong máy tính quan sát một đầu vào mới và cho ra kết quả đầu ra.
* Học không giám sát: Máy tính học từ dữ liệu đầu vào mà không có đầu ra, sau đó nó tự phát triển kiến thức về dữ liệu .
* Học bán giám sát: Cách học của học bán giám sát là sử dụng dữ liệu có nhãn và dự đoán từ mô hình ban đầu để tạo dữ liệu có nhãn giả, sau đó huấn luyện lại mô hình để cải thiện hiệu suất.
* Học tăng cương: Máy tính đưa ra hành động và nhận kết quả phản hồi từ môi trường. Sau đó máy tính học từ kết quả của hành động nó thực hiện.

**Ứng dụng của học máy:**

Có rất nhiều ứng dụng thực tế khác nhau của học máy. Hai lĩnh vực ứng dụng lớn nhất của học máy là khai phá dữ liệu (data mining) và nhận dạng mẫu (pattern recognition).

- KHAI PHÁ DỮ LIỆU là ứng dụng kỹ thuật học máy vào các cơ sở dữ liệu hoặc các tập dữ liệu lớn để phát hiện quy luật hay tri thức trong dữ liệu đó hoặc để dự đoán các thông tin quan tâm trong tương lai.

- NHẬN DẠNG MẪU là ứng dụng các kỹ thuật học máy để phát hiện các mẫu có tính quy luật trong dữ liệu, thường là dữ liệu hình ảnh, âm thanh. Bài toán nhận dạng mẫu cụ thể thường là xác định nhãn cho đầu vào cụ thể, ví dụ cho ảnh chụp mặt người, cần xác định đó là ai.

\* Xử lý ngôn ngữ tự nhiên (Natural Language Processing): xử lý văn bản, giao tiếp người - máy,...

\* Nhận dạng (Pattern Recognition): nhận dạng tiếng nói, chữ viết tay, vân tay, thị giác máy (Computer Vision),...

\* Tìm kiếm (Search Engine)

\* Chẩn đoán trong y tế: phân tích ảnh X-quang, các hệ chuyên gia chẩn đoán tự động.

\* Tin sinh học: phân loại chuỗi gene, quá trình hình thành gene/protein

\* Vật lý: phân tích ảnh văn thiên, tác động giữa các hạt,...

\* Phát hiện gian lận tài chính (financial fraud): gian lận thẻ tín dụng

\* Phân tích thị trường chứng khoán (stock market analysis)

\* Chơi trò chơi: tự động chơi cờ, hành động của các nhân vật ảo

Röbốt: là tông hợp của rất nhiều ngành khoa học, trong đó học máy tạo nên hệ thần kinh/bộ não của người máy.

**Lý do cần tới học máy:**

Học máy là một nhánh nghiên cứu rất quan trọng của trí tuệ nhân tạo với khá nhiều ứng dụng thành công trong thực tế. Có một số lý do giải thích cho sự cần thiết và phát triển của học máy:

+ Thứ nhất, rất khó xây dựng hệ thống thông minh có thể thực hiện các công việc liên quan đến trí tuệ như thị giác máy, xử lý ngôn ngữ tự nhiên mà không sử dụng tới kinh nghiệm và quá trình học. Thông thường, khi viết chương trình, cần có thuật toán rõ ràng để chuyển đổi đầu vào thành đầu ra. Tuy nhiên, trong nhiều bài toán, rất khó để xây dựng được thuật toán như vậy.

+ Thứ hai, nhiều ứng dụng đòi hỏi chương trình máy tính phải có khả năng thích nghi. Ví dụ, hành vi mua sắm của khách hàng có thể thay đổi theo thời điểm cụ thể trong ngày, trong năm, hoặc theo tuổi tác. Việc xây dựng thuật toán cố định cho những ứng dụng cần thích nghi và thay đổi là không phù hợp. Học máy mang lại khả năng thích nghi nhờ phân tích dữ liệu thu thập được.

+ Thứ ba, việc tìm được chuyên gia và thu thập được tri thức cần thiết cho việc thiết kế thuật toán để giải quyết các vấn đề tương đối khó, trong khi dữ liệu ngày càng nhiều và có thể thu thập dễ dàng hơn. Khả năng lưu trữ và tính toán của máy tính cũng ngày càng tăng, cho phép thực hiện thuật toán học máy trên dữ liệu có kích thước lớn.

+ Cuối cùng, khả năng học là một hoạt động trí tuệ quan trọng của con người, do vậy học máy luôn thu hút được sự quan tâm khi xây dựng hệ thống thông minh.