Họ tên sinh viên: Đỗ Duy Long

MSSV: 19110390

**Week 1 Exercirses**

1. Machine Learning là gì? So sánh nó với lập trình truyền thống.

Machine Learning là một phần của AI cung cấp cho các hệ thống khả năng học và cải thiện một cách tự động từ kinh nghiệm (data) mà không cần được lập trình một cách rõ ràng và chính xác. Machine Learning tập trung vào việc phát triển các chương trình có khả năng truy cập, xử lý và sử dụng data để học.

Đối với lập trình truyền thống, là quá trình thủ công mà người lập trình sẽ tạo nên các logic cũng như là rules của chương trình để giải quyết vấn đề.



Ngược lại, đối với Machine Learning thì là quá trình tìm kiếm tự động. Là một phương pháp tối ưu hóa nhằm tìm ra thuật toán thích hợp để giải quyết vấn đề dựa vào data có được.



2. Labeled data là gì? Nêu tên các bài toán thuộc loại Supervised learning và cho ví dụ cụ thể (khác với ví dụ trong bài giảng). Giải thích tại sao đó là Supervised learning.

Labeled data là data được gán nhãn (gán thêm một hoặc nhiều ý nghĩa và thông tin) để nhận diện data đó (text, audio, video, image, image có xe, email spam, ...).

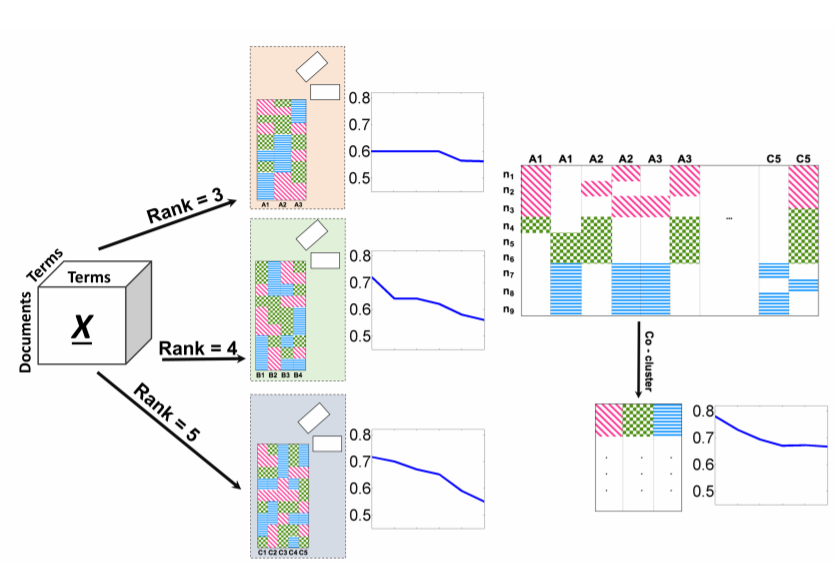
Các bài toán thuộc Supervised learning:

* Classification: phân loại audio (music audio, fx audio, impact fx audio, ...), phân loại dựa trên audio label nên thuộc Supervise learning.
* Regression: dự đoán giá vàng dựa trên giá trị của đồng USD, nguồn cung. Dự đoán dựa trên label là vàng và các features là đồng USD, nguồn cung nên thuộc Supervise learning.

3. Nêu tên các bài toán thuộc loại Unsupervised learning và cho ví dụ cụ thể (khác với ví dụ trong bài giảng). Giải thích tại sao đó là Unsupervised learning.

Các bài toán thuộc Unsupervise learning:

* Clustering: nhận biết fake news dựa trên nội dung bài báo.



* Anomaly detection: cảm biến của điện thoại (mobile sensor).
* Dimensionality reduction: lựa chọn và giảm features của data. Ví dụ để dự đoán lượng tiền khách hàng sẽ chi để mua đồ thì không cần quá cụ thể các món đồ đã mua mà cần lược bỏ bớt các features đó.



* Association rule learning: chuẩn đoán bệnh trong y tế, đề xuất video trên youtube.

4. Semi-supervised learning là gì? Cho ví dụ về Semi-supervised learning.

Semi-supervied learning là bài toán kết hợp giữa Supervised learning và Unsupervise learning, liên quan đển một lượng nhỏ các Labeled data và một lượng lớn các Unlabeled data.

Ví dụ: phân loại text. Semi-supervisor cho phép thuật toán học từ một lượng nhỏ labeled text trong khi vẫn phân loại một lượng lớn unlabeled text trong data.

5. Batch learning là gì? Online learning là gì? Cho ví dụ mỗi loại.

Batch learning là mô hình training của Machine learning mà lượng data được tích lũy trong một khoảng thời gian, sau đó được được train trong một khoảng thời gian trước khi được sử dụng. Hệ thống không có khả năng học thêm data, nếu muốn sử dụng lượng data mới thì cần phải retrain lại.

Online learning là mô hình được train trong quá trình sử dụng với lượng data được thay đổi liên tục. Ví dụ: recommend video youtube, nếu ta thay đổi loại video xem thường xuyên, recommend video cũng thay đổi.

6. Loại ML nào có thể dùng để huấn luyện robot dọn dẹp nhà cửa? Lý giải lựa chọn của bạn.

Chọn d) Kết hợp các loại trên. Kết hợp giữa Supervised learning và Reinforcement learning. Supervised learning hỗ trợ dự đoán khu vực có bụi để thực hiện dọn dẹp, dự đoán chuyển động sử dụng để thực hiện nó. Còn Reinforcement learning hỗ trợ khả năng học hành động dọn dẹp mới cũng như thay đổi phù hợp với không gian cần dọn dẹp.

7. Loại ML nào có thể dùng để nhận dạng virus máy tính? Lý giải lựa chọn của bạn.

Chọn d) Kết hợp các loại. Kết hợp giữa Supervised learning và Unsupervised learning. Supervised learning hỗ trợ nhận dạng vius dựa trên labeled data còn Unsupervise learning sẽ nhận dạng virus unlabeled (virus mới, chưa có nhãn).