[ML – 01] INTRODUCTION TO MACHINE LEARNING

Thời gian gần đây chúng ta được nghe rất nhiều về lĩnh vực này, Machine Learning và AI được nhắc đến không ngừng qua những sản phẩm "thông minh" của những "ông lớn" như Google, Apple, Tesla hay Microsoft… Hôm nay chúng ta sẽ bắt đầu loạt bài về Machine Learning (ML) nhằm giúp các bạn hiểu và có cái nhìn đúng đắn về nó đồng thời tạo nền tảng cơ bản cho những ai yêu thích có thể tìm tòi, học hỏi và phát triển ML nói riêng và AI nói chung.

1. Giới thiệu:

Hiện nay có rất nhiều sản phẩm đã được ứng dụng ML, nhiều trong số đó chúng ta sử dụng hàng ngày, tuy nhiên có thể ta chẳng biết thực ra chúng thông minh như thế nào khi sử dụng ML. Chẳng hạn :

– Google Translate: nếu như bạn hay sử dụng từ điển thì ứng dụng này quá đỗi quen thuộc, mặc dù đôi khi chúng ta thấy nó dịch mọi thứ hơi "ngô nghê" nhưng thực tế Google Translate được ứng dụng ML để hiểu được ngữ nghĩa của câu, cũng như nhận diện âm thanh và hình ảnh trong quá trình dịch.

– Google Photo: ứng dụng này có khả năng phân loại ảnh của ta thành từng sự kiện nhỏ. Khả năng này có được nhờ việc học từ hàng triệu bức ảnh trên internet.

– Facebook: khả năng nhận diện khuôn mặt để match với tài khoản facebook tương ứng hoạt động khá tốt. Bên cạnh đó facebook cũng đưa ra trợ lý ảo Facebook – M trong Facebook Messenger.

– iMessage: là ứng dụng tin nhắn của Apple được ứng dụng ML để hiểu ngữ nghĩa của câu trong tin nhắn và đem lại những đề nghị hỗ trợ người dùng đưa ra quyết định.

– Và rất nhiều những ứng dụng khác nữa…

2. Ý tưởng:

Trong khoa học máy tính (Computer Science) cũng như trong Toán học, kiến thức đưa chúng ta đi xa hơn không dừng lại ở những công thức hay cách thức giải quyết vấn đề mà là những ý tưởng. Vậy ý tưởng trong ML là gì ?

* Về khía cạnh Trí tuệ nhân tạo: ML là kĩ thuật giúp chúng ta lập trình những thứ không thể lập trình một cách rõ ràng.

* Về mặt dữ liệu: Với sự bùng nổ về số lượng cũng chất lượng của dữ liệu hiện nay, câu hỏi đặt ra cho tất cả những lĩnh vực có liên quan tới dữ liệu là chũng ta sẽ có được gì từ chúng (không chỉ để thống kê, lưu trữ…). Câu trả lời là ta sẽ học được gì đó từ chúng.

* Nếu như các bạn từng xem bộ phim gần đây của Adam Sandler với tựa đề "Pixel". Phim có nói về một cậu bé có khả năng chơi game vượt trội nhờ nhận ra được những pattern trong các game, và trong thực tế cũng đúng như vậy, mọi thứ đều có khuôn mẫu cụ thể nào đó, hoặc chí ít nó sẽ "hao hao" với những khuôn mẫu chúng ta có thể nghĩ ra được.

* Vậy nên ML là một nhánh của AI giúp cho máy tính có khả năng học hỏi mà không cần phải lập trình một cách cụ thể, rõ ràng.

Screen Shot 2016-07-23 at 6.52.44 AM

Con người nhận thông tin, xử lý trong não bộ và đưa ra giải pháp. ML cũng cố gắng bắt chước quá trình này thông qua việc xây dựng hàm số f(x) với x là input và y = f(x) là output.

3. Một chút về lịch sử:

ML không phải là một lĩnh vực mới, nhiều ứng dụng của chúng thời nay vẫn chỉ là ứng dụng những kết quả nghiên cứu trước đây.

Lần đầu tiên ML xuất hiện là vào năm 1952, nó được ứng dụng vào chương trình chơi cờ Đam (game of checker) của Arther Samuel.

Đến năm 1957, mạng neuron nhân tạo đầu tiên được thiết kế (với tên gọi perceptron).

Năm 1967, giải thuật "Nearest neighbor" được phát minh, tiền đề cho nhiều giải thuật sau này.

4. Máy học như thế nào ?

Screen Shot 2016-07-23 at 7.05.47 AM

Như đã nói ở trên, mục tiêu của chúng ta là cố gắng xây dựng hàm số f(x) để xứ lý input và đưa ra output tương tự như khi con người làm. Tuy nhiên cách làm việc của bộ não con người tới giờ vẫn còn khá nhiều ẩn số, nên dù mục tiêu là vậy nhưng chúng ta vẫn chỉ đang xây dựng hàm h(x) (hypothesis) mang tính giả thuyết. Mục tiêu bây giờ trở thành sao cho h(x) có kết quả giống f(x) nhất.

Mặc dù f(x) ta không xác định được nhưng chính dữ liệu sử dụng cho máy học hỏi lại là những bộ (input, output) của f(x). Vậy nên đơn thuần ta sẽ chỉ phải làm h(x) giống f(x) nhất có thể trên tập dữ liệu sử dụng cho việc học của máy.

=> Vấn để mới được đặt ra: làm sao để h(x) giống với f(x) nhất ?

Lúc này chúng ta cần phải xác định sự sai khác giữa h(x) và f(x), sau đó giảm thiểu sự sai khác này xuống mức cực tiểu. Khi đó h(x) sẽ gần với f(x) nhất trên mô hình ta đưa ra. Sự sai khác đó thường được biểu diễn bằng hàm số J(θ) trong đó θ là những tham số ta đưa vào h(x) và có thể thay đổi được. (Chính nhờ những tham số có thể thay đổi được này máy mới có khả năng học hỏi).

Kết luật: Với mỗi vấn đề cụ thể, chúng ta xây dựng một mô hình được mô ta thông qua hàm giả thuyết h(x) trong đó có chứa những tham số θ có thể thay đổi. Bên cạnh đó chúng ta xác định một hàm sai khác J(θ) (hay còn được gọi là cost function hay error function). Quá trình học của máy chính là quá trình chúng ta tối thiểu hóa J(θ).

5. Tổng kết:

Đọc đến đây hẳn các bạn đã mù mờ về công việc mà ML đang làm hàng ngày. Đây không phải là một lĩnh vực mới, nhưng cũng không cũ, và bây giờ nó đang phát triển mạnh mẽ hơn bao giờ hết. Trong bài viết sau tôi giúp các bạn tìm hiểu sâu hơn về cách ML tối thiểu hoá J(θ). Cảm ơn sự theo dõi của các bạn và hẹn gặp lại trong bài viết sau.