Học máy

Xếp hàng các trang web khi search, đề xuất phim, email tự nhận diện thư rác

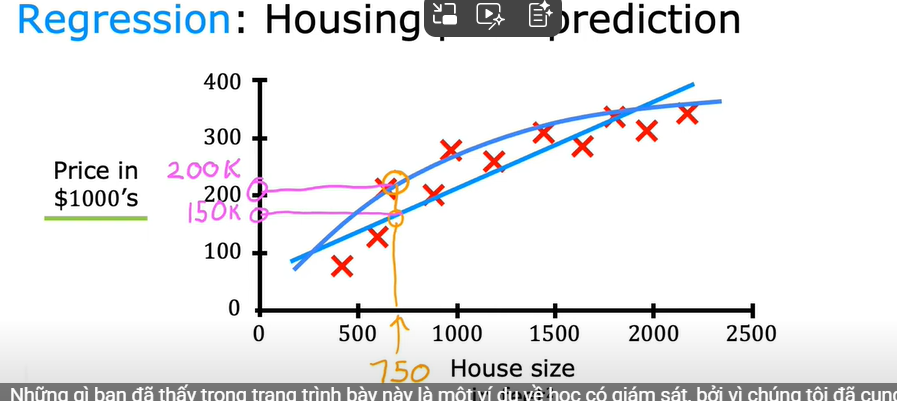
Tối ưu sản xuất điện gió, chẩn đoán chỉnh xác trong y tế. Khoa học khiến máy tình học mà không cần lập trình rõ ràng

Gồm : Học có giám sát(used most) và học không có giám sát, hệ thống đề xuất

Học có giám sát: liên quan đến thuật toán x->y, ánh xạ đầu cuối, cho máy các ví dụ có thông tin đầu cuối, sau đó cuối cùng dựa vào lượng lớn số ví dụ khác nhau đó mà đưa ra xác suất, dự đoán chính xác kết quả cuối khi chỉ cần đầu vào x. Có thể sẽ móc nối các điểm kết quả thành 1 đường thẳng hoặc đường cong(dựa trên thuật toán).

**Học cách dự đoán các con số trong vô số các con số có thể.**

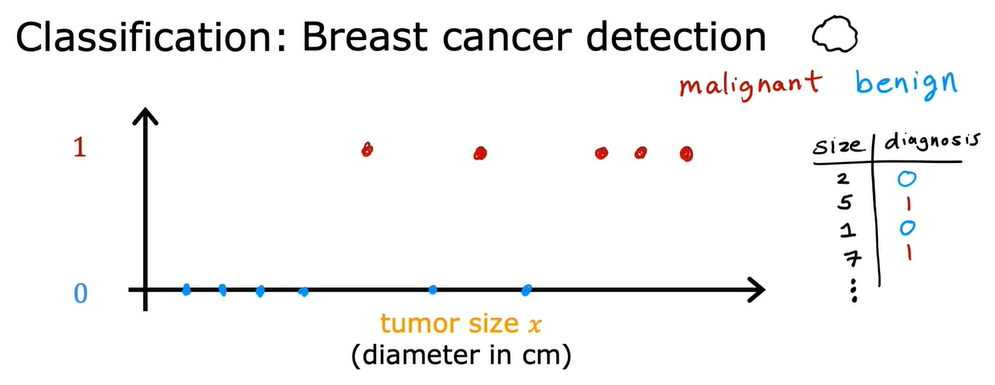
?! Làm thế nào để có thuật toán để lựa chọn 1 cách có hệ thống đường thằng, đường cong hoặc thứ khác phù hợp nhất để khớp với 1 tệp dữ liệu?



Loại dự đoán này là 1 loại hình học có giám sát cụ thể được gọi là **hồi quy**

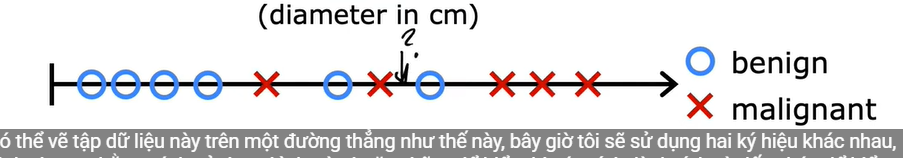
**Thuật toán phân loại**  - Phát hiện ung thư vú

Đầu vào: size cục u -> f(x) : lành tính(0), ác tính(1)



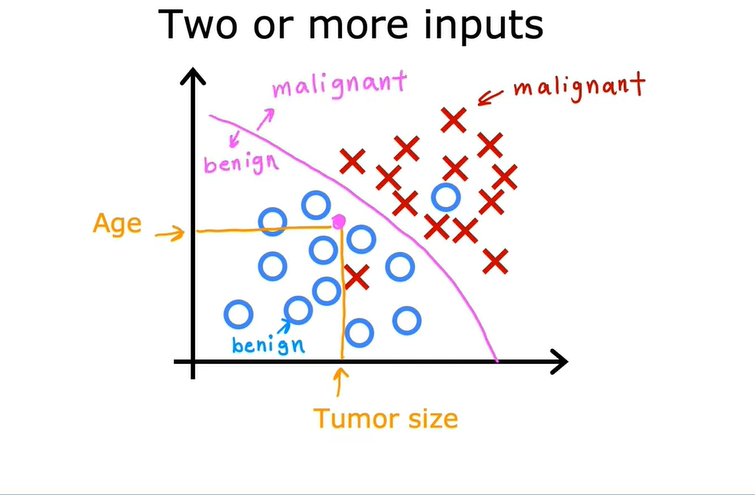
Điều khác với hồi quy là **đang cố gắng dự đoán chỉ 1 số lượng nhỏ các đầu ra** hoặc danh mục có thể

(Hồi quy : cố gắng dự đoán bất kỳ số nào trong vô số các số có thể)



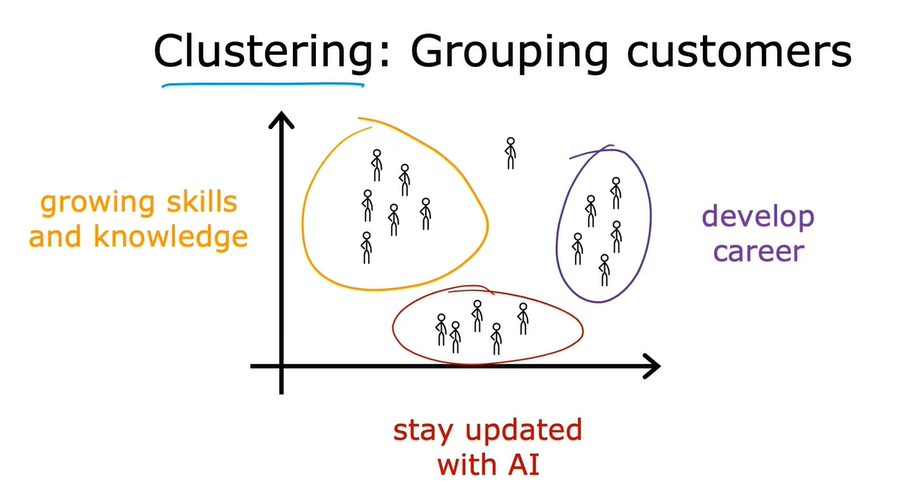
Khác với hồi quy, khi diễn giải các con số là phân loại dự đoán 1 tập hợp hữu hạn giới hạn các đầu mục có thể như 0, 1 hoặc 2 nhưng không phải tất cả các số có thể ở giữa như 1,7; 1,2;...

Có thể sử dụng nhiều hơn 1 loại giá trị đầu vào để dự đoán đầu ra(two or more inputs)

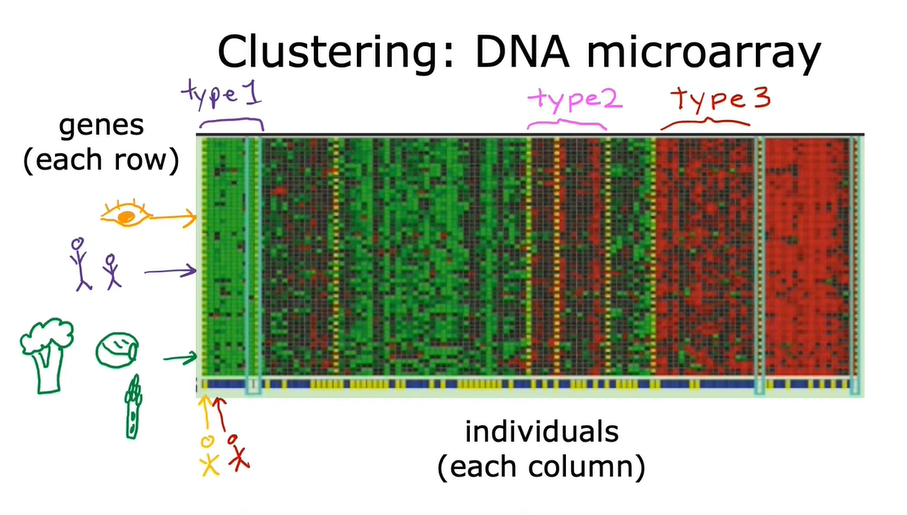


**Học không giám sát : Tìm cấu trúc trong dữ liệu**

Clustring(Thuật toán phân cụm) : lấy dữ liệu không có nhãn và cố gắng nhóm chúng thành các cụm







Học máy không giám sát : dữ liệu chỉ đi kèm với đầu vào x nhưng không có đầu ra y và **thuật toán phải tìm 1 số cấu trúc hoặc 1 số mẫu hoặc 1 cái gì đó thú vị** trong dữ liệu.

**Anomaly detection(phát hiện bất thường)** : phát hiện gian lận trong giao dịnh hay điều gì đó bất thường.

**Dimensionnality reduction(giảm chiều dữ liệu)** : nén 1 tập dữ liệu lớn thành 1 tập dữ liệu nhỏ hơn nhiều trong khi mất ít thông tin nhất có thể.