



CANTHO UNIVERSITY

# Máy học nâng cao Chương 1

Phạm Nguyên Khang  
[pnkhang@cit.ctu.edu.vn](mailto:pnkhang@cit.ctu.edu.vn)



CANTHO UNIVERSITY

# Tại sao phải học NLMH

- Tiến bộ gần đây trong giải thuật và tin học lý thuyết
- Dữ liệu ngày càng nhiều hơn, nhất là dữ liệu trực tuyến
- Năng lực tính toán của máy tính ngày càng phát triển (phần cứng)
- Đất dụng vô dành cho dân máy học
  - Khai khoáng dữ liệu (data mining): sử dụng dữ liệu quá khứ để dự đoán tương lai → cải thiện quá trình quyết định.
    - Thông tin y khoa → kiến thức y khoa
  - Ứng dụng phần mềm mà ta không thể lập trình thủ công
    - Xe tự hành
    - Nhận dạng tiếng nói
  - Chương trình tự tuỳ biến
    - Chương trình đọc tin tức có thể học để biết được sự quan tâm của người dùng.



CANTHO UNIVERSITY

# Nhiệm vụ tiêu biểu của DM

- Cho dữ liệu:
  - 9714 hồ sơ bệnh nhân, mỗi hồ sơ dành cho một phụ nữ mang thai và sinh con
  - Mỗi hồ sơ có 215 đặc điểm liên quan đến bệnh nhân: độ tuổi, mang thai lần đầu, có mắc bệnh tiểu đường không, ...
- Học để dự đoán:
  - Bệnh nhân nào có nguy cơ bị mắc bệnh cao dựa vào các đặc điểm của họ.



# Nhiệm vụ tiêu biểu của DM

Customer103: (time=t0)

Years of credit: 9  
Loan balance: \$2,400  
Income: \$52k  
Own House: Yes  
Other delinquent accts: 2  
Max billing cycles late: 3  
Profitable customer?: ?  
...

Customer103: (time=t1)

Years of credit: 9  
Loan balance: \$3,250  
Income: ?  
Own House: Yes  
Other delinquent accts: 2  
Max billing cycles late: 4  
Profitable customer?: ?  
...

...

Customer103: (time=tn)

Years of credit: 9  
Loan balance: \$4,500  
Income: ?  
Own House: Yes  
Other delinquent accts: 3  
Max billing cycles late: 6  
**Profitable customer?: No**  
...

- Các luật được học từ dữ liệu trên là:
- If Other\_Delinquent-Accounts > 2 and Number-Delinquent-Billing-Cycles > 1

then

Profitable-Customer = No



CANTHO UNIVERSITY

# Nhiệm vụ tiêu biểu của DM

## Customer purchase behavior:

*Customer103: (time=t0)*

Sex: M  
Age: 53  
Income: \$50k  
Own House: Yes  
MS Products: Word  
Computer: 386 PC  
Purchase Excel?: ?  
...

*Customer103: (time=t1)*

Sex: M  
Age: 53  
Income: \$50k  
Own House: Yes  
MS Products: Word  
Computer: Pentium  
Purchase Excel?: ?  
...

*Customer103: (time=tn)*

Sex: M  
Age: 53  
Income: \$50k  
Own House: Yes  
MS Products: Word  
Computer: Pentium  
**Purchase Excel?: Yes**  
...



CANTHO UNIVERSITY

# Nhiệm vụ tiêu biểu của DM

## Customer retention:

*Customer103: (time=t0)*

Sex: M  
Age: 53  
Income: \$50k  
Own House: Yes  
Checking: \$5k  
Savings: \$15k  
Current-customer?: yes

*Customer103: (time=t1)*

Sex: M  
Age: 53  
Income: \$50k  
Own House: Yes  
Checking: \$20k  
Savings: \$0  
Current-customer?: yes

...

*Customer103: (time=tn)*

Sex: M  
Age: 53  
Income: \$50k  
Own House: Yes  
Checking: \$0  
Savings: \$0  
**Current-customer?: No**



CANTHO UNIVERSITY

# Nhiệm vụ tiêu biểu của DM

## Customer retention:

*Customer103: (time=t0)*

Sex: M  
Age: 53  
Income: \$50k  
Own House: Yes  
Checking: \$5k  
Savings: \$15k  
Current-customer?: yes

*Customer103: (time=t1)*

Sex: M  
Age: 53  
Income: \$50k  
Own House: Yes  
Checking: \$20k  
Savings: \$0  
Current-customer?: yes

...

*Customer103: (time=tn)*

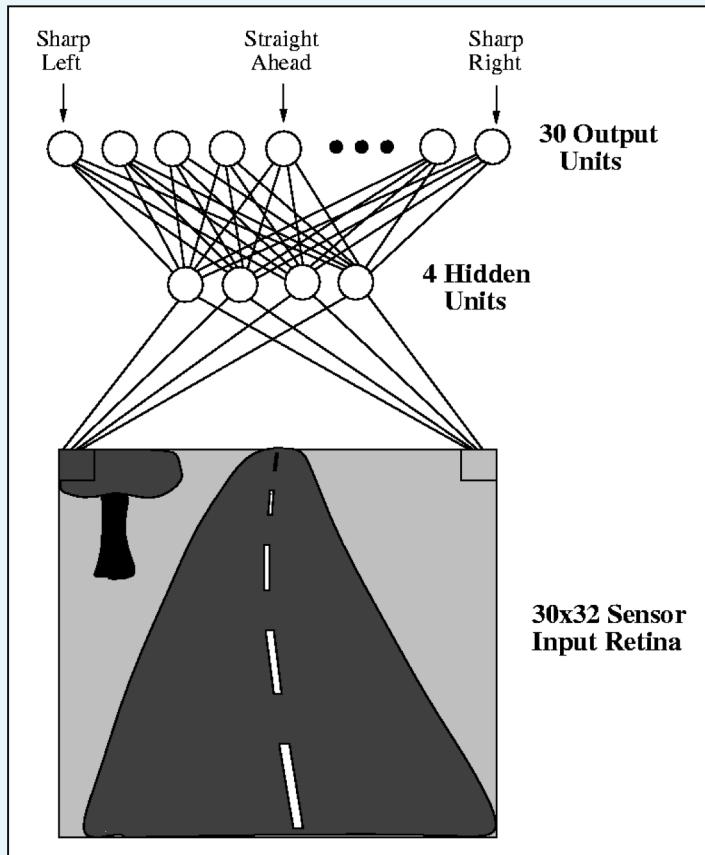
Sex: M  
Age: 53  
Income: \$50k  
Own House: Yes  
Checking: \$0  
Savings: \$0  
**Current-customer?: No**



CANTHO UNIVERSITY

# Vấn đề khó lập trình thủ công

- Chương trình ALVINN có thể tự lái xe với tốc độ 70mph (112.654 km/h) trên đường cao tốc





CANTHO UNIVERSITY

# Phần mềm tự tuỳ biến cho người dùng

- Phần mềm tự thay đổi điều chỉnh giao diện tùy theo người dùng
- Trang web tự đề nghị nội dung cho người đọc
- Trang web bán hàng tự đề nghị mặt hàng cho người mua
- ...



CANTHO UNIVERSITY

## Các môn học khác có liên quan

- Trí tuệ nhân tạo
- Phương pháp Bayes
- Lý thuyết độ phức tạp tính toán
- Lý thuyết điều khiển
- Lý thuyết thông tin
- Triết học
- Tâm lý học và thần kinh học (neurobiology)
- Thống kê
- ...



CANTHO UNIVERSITY

# Học là gì ?

- Học (Mitchel)= Cải thiện tác vụ (task) nào đó bằng kinh nghiệm
  - Cải thiện tác vụ T,
  - Với độ đo hiệu quả P
  - Dựa trên kinh nghiệm E
- Ví dụ: học chơi cờ
  - T: chơi cờ
  - P: % ván thắng
  - E: cơ hội để tự đánh với chính mình



CANTHO UNIVERSITY

# Học chơi cờ

- T: chơi cờ
- P: % ván thắng
- Kinh nghiệm gì ?
- Nên học gì ?
- Biểu diễn nó như thế nào ?
- Giải thuật nào để học ?



CANTHO UNIVERSITY

# Kiểu dạy/huấn luyện

- Trực tiếp hay gián tiếp ?
- Có thầy hay không có thầy ?
- Vấn đề: việc huấn luyện có phải là kinh nghiệm có thể biểu diễn được của mục tiêu hiệu quả ?



CANTHO UNIVERSITY

# Chọn hàm mục tiêu

- Lựa chọn nước đi: Bàn cờ → Nước đi ??
- V: Bàn cờ → R ?? Lượng giá bàn cờ
- ...



CANTHO UNIVERSITY

# Dự báo giá đất dựa trên diện tích

- Tác vụ T: ???
- Độ đo hiệu quả P: ???
- Kinh nghiệm E: ???

Living area (feet <sup>2</sup> )	Price (1000\$s)
2104	400
1600	330
2400	369
1416	232
3000	540
:	:

Mảnh đất có dt 2500 ft<sup>2</sup> sẽ có giá bằng bao nhiêu ?



CANTHO UNIVERSITY

## Ví dụ khác

- Nhiệm vụ:
- Độ đo hiệu quả:
- Kinh nghiệm:



CANTHO UNIVERSITY

# Phân loại học máy

- Học có giám sát (có thầy)
  - Phân lớp
  - Hồi quy
- Học không có giám sát (không có thầy)
  - Gom cụm
  - Ước lượng hàm mật độ xác suất
- Học bán giám sát
  - Gom cụm + gán nhãn => Phân lớp