

# Machine Learning

Bộ môn Khoa học dữ liệu  
Khoa Công nghệ thông tin  
Trường Đại học Công nghiệp thành phố Hồ Chí Minh-IUH

*Các bài tập từ 2-7 làm bằng 2 cách:*

- *Tự tính bằng tay*
- *Hướng dẫn cho máy tính tính và in ra kết quả*

## **Bài 1:**

1- Khởi tạo các ma trận tự động theo các yêu cầu sau:

- Ma trận  $A[3,4]$  toàn giá trị 0
- Ma trận  $B[2,4]$  toàn giá trị 1
- Ma trận  $C[4,4]$  toàn giá trị 5
- Ma trận  $D[5,5]$  với các giá trị trên đường chéo chính là 1
- Ma trận  $E[4,6]$  với các giá trị ngẫu nhiên

2- Khởi tạo các ma trận vuông  $A[5,5]$  toàn giá trị 1 ở đường chéo chính, 0 ở các ô còn lại một cách tự động. Khởi tạo ma trận vuông  $B[5,5]$  toàn giá trị 1 ở các ô là tổng của  $2*i+j$  (hàng, cột), 0 ở các ô còn lại một cách tự động. Tính bằng tay và tính tự động các yêu cầu sau:

- Tính định thức của A sử dụng thư viện
- Tính định thức của B sử dụng thư viện
- Tính  $A*B$
- Tính  $A+B$
- Tính  $A-B$
- Tính  $4*A/B$

## **Bài 2:**

Một sọt Cam có 10 trái trong đó có 4 trái hư. Lấy ngẫu nhiên ra ba trái.

- Tính xác suất lấy được 3 trái hư.
- Tính xác suất lấy được 1 trái hư.
- Tính xác suất lấy được ít nhất một trái hư.
- Tính xác suất lấy được nhiều nhất 2 trái hư.

### **Bài 3:**

Một gia đình có 10 người con. Giả sử xác suất sinh con trai, con gái như nhau. Tính xác suất:

- a) Không có con trai.
- b) Có 5 con trai và 5 con gái.
- c) Số trai từ 5 đến 7.

### **Bài 4:**

a. Một nhóm học sinh có 10 em, trong đó có 3 em học loại giỏi, 4 loại khá, còn lại là trung bình. Từ nhóm đó chọn ngẫu nhiên ra 3 học sinh. Gọi X là số học sinh giỏi trong số học sinh chọn ra

- Lập bảng phân phối xác suất của X, đưa dữ liệu bảng này vào file data.txt

Viết các hàm tính Kỳ vọng, Phương sai, Độ lệch chuẩn và Áp dụng các hàm đã viết cho dữ liệu biến ngẫu nhiên rời rạc được cung cấp ở file data.txt

b. Áp dụng các hàm đã viết tính Kỳ vọng, Phương sai, Độ lệch chuẩn cho dữ liệu biến ngẫu nhiên liên tục như mô tả sau:

Nhu cầu hàng năm về loại hàng hóa A là **biến ngẫu nhiên liên tục X có hàm mật độ** xác suất như sau: (đơn vị: ngàn sản phẩm)

$$f(x) = \begin{cases} k(30-x) & \text{với } x \in [0,30] \\ 0 & x \notin [0,30] \end{cases}$$

c. Sử dụng thư viện có sẵn tính Kỳ vọng, Phương sai, Độ lệch chuẩn cho dữ liệu ở câu a

d. Sử dụng thư viện có sẵn tính Kỳ vọng, Phương sai, Độ lệch chuẩn cho dữ liệu ở câu b

### **Bài 5:**

Những sự ảnh hưởng theo độ tuổi không chỉ xảy ra ở điều mà bạn xem thấy trên tivi mà còn ở nơi mà bạn xem tivi. Một nghiên cứu (Darnay, 1994, trang 784) đã cho thấy rằng những người Mỹ lớn tuổi hơn thì ít xem tivi trên giường ngủ hơn so với

những người trẻ tuổi và thường xem tivi tại phòng ăn nhiều hơn. Với dữ liệu đã cho trong bảng, giả định rằng cỡ mẫu cho từng nhóm tuổi là 100.

Khu vực	25 đến 44	45 đến 69	60 trở lên
Phòng khách/phòng sinh hoạt chung/phòng làm việc	95%	95%	93%
Phòng ngủ	58%	57%	45%
Nhà bếp	12%	20%	20%
Phòng ăn	10%	10%	10%

- Tìm ước lượng khoảng tin cậy 95% về sự khác biệt trong các tỷ lệ của người Mỹ trong độ tuổi từ 45 đến 69 và những người từ 60 tuổi trở lên mà xem tivi tại phòng ăn.
- Ước lượng sự khác biệt giữa các tỷ lệ của người Mỹ trong nhóm tuổi từ 25 đến 59 và những người trong nhóm độ tuổi từ 60 trở lên mà xem tivi tại phòng khách, phòng sinh hoạt chung hay phòng riêng làm việc và tìm biên sai số ước lượng. [Gợi ý: Tỷ lệ cho nhóm tuổi từ 25 đến 59 sẽ là bình quân giản đơn của các tỷ lệ riêng lẻ dựa trên cỡ mẫu là 2000]

### **Bài 6:**

Một cuộc điều tra 100 đại lý mua hàng tạo ra một ước lượng về tỷ lệ những người bán buôn ống nước polyvinyl mà có kế hoạch gia tăng sự mua hàng của mình trong năm tới. Biên sai số, 0.096, là tương đối lớn. Giả định rằng tổ chức tiếp thị tiến hành cuộc điều tra này được yêu cầu phải thực hiện một cuộc điều tra mới và đạt được một ước lượng chính xác trong giới hạn 0.04 với xác suất bằng với 0.90. Xấp xỉ có bao nhiêu nhà bán buôn ắt đã phải tính đến trong cuộc điều tra này?

### **Bài 7:**

Một giám đốc nhân sự mong muốn so sánh tính hiệu quả của hai phương pháp huấn luyện các nhân viên công nghiệp nhằm thực hiện một hoạt động lắp ráp nào đó. Một số lượng nhân viên được chia thành hai nhóm bằng nhau, nhóm thứ nhất nhận được phương pháp huấn luyện 1 và nhóm thứ hai được huấn luyện bằng phương pháp 2. Mỗi nhóm sẽ thực hiện hoạt động lắp ráp này, và độ dài của thời gian lắp ráp sẽ được ghi nhận. Người ta kỳ vọng rằng các đại lượng cho cả hai nhóm sẽ có một khoảng xấp xỉ 8 phút. Để cho ước lượng về sự khác biệt về thời gian trung bình để lắp ráp chính xác trong giới hạn 1 phút với xác suất bằng với 0.95, thì cần phải đưa bao nhiêu công nhân vào mỗi nhóm huấn luyện?