

Mô hình hồi quy Logistic

NHÓM 3

Lai Đức Thắng - 20163830

GVHD : TS. TRẦN MẠNH TUẤN

Viện Toán ứng dụng và Tin học, ĐHBK Hà Nội

Hà Nội, tháng 07 năm 2020

Tổng quan

- Vĩnh thức
- Hình vuông Latin
- Phân tích hồi quy
- Giới thiệu

Vĩnh thức

Giả sử $M = (m_{ij})$ là ma trận số thức $n \times n$. Khi đó vĩnh thức (Permanents) của ma trận M (kí hiệu là per):

$$\text{per} M = \sum_{\sigma \in S_n} m_{1\sigma(1)} m_{2\sigma(2)} \dots m_{n\sigma(n)}$$

trong đó S_n là tập tất cả các hoán vị của $\{1, 2, \dots, n\}$.

Khái niệm

1	2	3	4
2	1	4	3
4	3	1	2
3	4	2	1

Table: Hình vuông Latin bậc 4.

Bài toán

Xét bài toán: Xác định số hình vuông latin $L(n)$ bậc n . Xét một vài ví dụ nhỏ :

$$n = 1: \begin{array}{|c|} \hline 1 \\ \hline \end{array}: L(1) = 1$$

$$n = 2: \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 2 & 1 \\ \hline \end{array} \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 1 & 2 \\ \hline \end{array}: L(2) = 2$$

$$n = 3: \begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 2 & 3 & 1 \\ \hline 3 & 1 & 2 \\ \hline \end{array}: L(3) = 12$$

$$L(1) = 1, L(2) = 2, L(3) = 12, L(4) = 576, L(5) = 161280$$

Bài toán

Xét bài toán: Xác định số hình vuông latin $L(n)$ bậc n . Xét một vài ví dụ nhỏ :
 $n = 1$:

1

: $L(1) = 1$

$n = 2$:

1	2
2	1

2	1
1	2

: $L(2) = 2$

$n = 3$:

1	2	3
2	3	1
3	1	2

: $L(3) = 12$

$$L(1) = 1, L(2) = 2, L(3) = 12, L(4) = 576, L(5) = 161280$$

$$? \leq L(n) \leq ?$$

Mô hình bài toán

Xét một hình vuông $n \times n$ và điền vào đó các số $1, 2, \dots, n$ sao cho đó là một hình vuông Latin

2	4	1	3

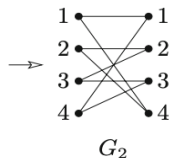
Mô hình bài toán

Xét một đồ thị hai phía $G_k = (U \cup V, E)$, trong đó U là tập các phần tử $\{1, 2, \dots, n\}$ và V là tập các vị trí của cột, sao cho:

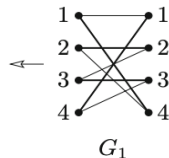
$ij \in E \Leftrightarrow i$ không xuất hiện trong cột thứ j của R .

2	4	1	3
3	1	2	4

R



2	4	1	3
3	1	2	4
4	2	3	1



Phân tích hồi quy

- **Phân tích hồi quy** nghiên cứu sự phụ thuộc của một biến trả lời (response variable) vào một hoặc nhiều biến dự báo (predictors).
 - (x_1, x_2, \dots, x_D) : các biến dự báo (là các thành phần xác định được)
 - y : biến trả lời
- Có 2 dạng phân tích hồi quy:
 - Hồi quy tuyến tính
 - Hồi quy phi tuyến

Hồi quy tuyến tính

Hàm dự báo $h_{\theta}(\mathbf{x})$ được xấp xỉ bởi một hàm tuyến tính của \mathbf{x} .

$$h_{\theta}(\mathbf{x}) = \theta_0 + \theta_1 x_1 + \cdots + \theta_D x_D$$

Nếu bổ sung thêm đặc trưng cố định $x_0 \equiv 1$ thì ta có thể biểu diễn h dưới dạng

$$h_{\theta}(\mathbf{x}) = \sum_{j=0}^D \theta_j x_j = \theta^T \mathbf{x}.$$

Giới thiệu

- 1 Mô hình hồi quy logistic được sử dụng rộng rãi trong nhiều bài toán thống kê và học máy
- 2 Là một mô hình hồi quy nhằm dự đoán giá trị đầu ra rời rạc (discrete target variable) ứng với một véc tơ đầu vào. Việc này tương đương với chuyện phân loại các đầu vào x vào các nhóm y tương ứng (Bài toán phân loại)
- 3 Ví dụ: một bức ảnh có chứa một con mèo hay không. Thì ở đây ta coi đầu vào là các pixel của bức ảnh, đầu ra $y=1$ nếu bức ảnh có một con mèo và $y=0$ nếu bức ảnh không có con mèo nào.
- 4 Để đơn giản, nhóm em đã tìm hiểu mô hình và cách giải quyết cho bài toán phân loại nhị phân tức là $y = \{0, 1\}$.

Ví dụ thực tế

Mô hình nghiên cứu mối quan hệ quả bệnh tim và độ tuổi. Trong bảng số liệu thống kê sau, biến $CD = 1$ tức là người đó có mắc bệnh tim và $CD = 0$ tức là người đó không mắc bệnh tim. Dựa vào dữ liệu thống kê lớn với đa dạng độ tuổi người ta tìm cách trả lời câu hỏi liệu có mối quan hệ nào giữa độ tuổi và khả năng bị bệnh tim của một người hay không?

Đây là một phần nhỏ trong dữ liệu thô của 100 người tham gia nghiên cứu.

Chúng ta có thể nhóm mỗi người trong ví dụ trên vào các nhóm tuổi và quan sát tỷ lệ phần trăm mắc bệnh tim.