

# PHÂN TÍCH HỒI QUY

## THỰC HÀNH PHƯƠNG PHÁP SỐ CHO KHOA HỌC DỮ LIỆU

Ngày 18 tháng 5 năm 2023

## Liên hệ

**GOOGLE CLASSROOM:** **mrhvcvm**

**TRỢ GIẢNG:**

- ▶ Nguyễn Thị Kiều Trang: ntktrang@hcmus.edu.vn
- ▶ Lý Như Bình: lnbinh@hcmus.edu.vn

**LƯU Ý:**

- ▶ Email đăng nhập google classroom thể hiện đầy đủ họ và tên, tránh sử dụng email có biệt danh.
- ▶ Tiêu đề mail (bắt buộc):  
[2023-HK2-THPPSKHDL] [Tiêu đề thư]  
VD: [2023-HK2-THPPSKHDL] HỎI BÀI  
Vui lòng giới thiệu họ tên, MSSV và tên ca học khi gửi email.

## Một vài điều về lớp

**Điểm thực hành:** Chiếm 40% tổng điểm:

- ▶ Điểm danh: 1 điểm (Mỗi buổi)
- ▶ Bài tập: 3 điểm (Nộp bài tập thực hành mỗi tuần)

**Cách thức nộp bài:**

- ▶ Nộp trên google classroom
- ▶ Nộp file .txt
- ▶ Tên file: Y\_MSSV\_Hoten\_baix.txt,
  - ▶  $Y = LB$  nếu bạn học phòng C204.
  - ▶  $Y = LT$  nếu bạn học phòng C203.
  - ▶  $x \in \{1, 2, 3, 4, \dots\}$

## Phân tích hồi quy tuyến tính

**Bài 1:** Cho bảng số liệu sau:

| STT | Diện tích<br>( $m^2$ ) | Số phòng<br>ngủ | Khoảng cách<br>tới TT | Giá<br>(tỷ VND) |
|-----|------------------------|-----------------|-----------------------|-----------------|
| 1   | 40                     | 1               | 30                    | 1.1             |
| 2   | 60                     | 2               | 32                    | 1.55            |
| 3   | 53                     | 2               | 30.1                  | 1.68            |
| 4   | 71                     | 2               | 35.7                  | 1.75            |
| 5   | 80                     | 2               | 24.5                  | 5.5             |
| 6   | 56                     | 2               | 27.6                  | 2.3             |
| 7   | 75                     | 2               | 27.6                  | 3               |
| 8   | 79                     | 2               | 27.6                  | 3.5             |
| 9   | 56                     | 2               | 29.7                  | 2.4             |

| <b>STT</b> | <b>Diện tích<br/>(<math>m^2</math>)</b> | <b>Số phòng<br/>ngủ</b> | <b>Khoảng cách<br/>tối TT</b> | <b>Giá<br/>(tỷ VND)</b> |
|------------|---|-------------------------|-------------------------------|-------------------------|
| 10         | 60                                      | 2                       | 29.7                          | 2.9                     |
| 11         | 72                                      | 2                       | 29.7                          | 3                       |
| 12         | 95                                      | 3                       | 29.7                          | 4.2                     |
| 13         | 47                                      | 1                       | 19.3                          | 1.5                     |
| 14         | 91                                      | 2                       | 18.1                          | 2.2                     |
| 15         | 68                                      | 1                       | 21.4                          | 1.5                     |
| 16         | 69                                      | 2                       | 17.5                          | 3.15                    |
| 17         | 82                                      | 2                       | 25.1                          | 3.4                     |
| 18         | 60                                      | 2                       | 26.5                          | 2.245                   |
| 19         | 68                                      | 2                       | 26.5                          | 2.4                     |

Dựa vào bảng số liệu trên, hãy dự đoán giá của một căn nhà có diện tích là  $79m^2$ , 2 phòng ngủ, khoảng cách tới trung tâm là 26.5 km bằng cách:

- a) Giải phương trình đạo hàm mất mát
- b) Dùng các thuật toán Gradient descent, Accelerated gradient descent, Stochastic gradient descent.
- c) Dùng thư viện scikit-learn

Biết giá trị thực tế của căn nhà trên là 2.5 tỷ VND, hãy so sánh các kết quả trên với nhau.

**Bài 2:** Cho bảng số liệu sau:

| <b>STT</b> | <b>Chiều cao (cm)</b> | <b>Cân nặng (kg)</b> |
|------------|-----------------------|----------------------|
| 1          | 147                   | 49                   |
| 2          | 150                   | 50                   |
| 3          | 153                   | 51                   |
| 4          | 155                   | 52                   |
| 5          | 158                   | 54                   |
| 6          | 160                   | 56                   |
| 7          | 163                   | 58                   |
| 8          | 168                   | 60                   |
| 9          | 170                   | 72                   |
| 10         | 173                   | 63                   |
| 11         | 175                   | 64                   |
| 12         | 178                   | 66                   |
| 13         | 180                   | 67                   |
| 14         | 183                   | 68                   |

Bài toán đặt ra là từ bảng số liệu trên, hãy dự đoán cân nặng của một người có chiều cao là  $165\text{ cm}$  bằng cách:

- a) Giải phương trình đạo hàm mất mát.
- b) Dùng các thuật toán Gradient descent, Accelerated gradient descent, Stochastic gradient descent.
- c) Sử dụng thư viện scikit-learn.

Biết cân nặng thực tế của người đó trên là  $59\text{ kg}$ , hãy so sánh các kết quả trên với nhau.



## Phân tích hồi quy logistic

**Bài 3:** Cho bảng số liệu sau:

|   | Lương | Thời gian làm việc | Cho vay |
|---|-------|--------------------|---------|
| 0 | 10    | 1.0                | 1       |
| 1 | 5     | 2.0                | 1       |
| 2 | 6     | 1.8                | 1       |
| 3 | 7     | 1.0                | 1       |
| 4 | 8     | 2.0                | 1       |
| 5 | 9     | 0.5                | 1       |
| 6 | 4     | 3.0                | 1       |
| 7 | 5     | 2.5                | 1       |
| 8 | 8     | 1.0                | 1       |
| 9 | 4     | 2.50               | 1       |

|    | Lương | Thời gian làm việc | Cho vay |
|----|-------|--------------------|---------|
| 10 | 8     | 0.10               | 0       |
| 11 | 7     | 0.15               | 0       |
| 12 | 4     | 1.00               | 0       |
| 13 | 5     | 0.80               | 0       |
| 14 | 7     | 0.30               | 0       |
| 15 | 4     | 1.00               | 0       |
| 16 | 5     | 0.50               | 0       |
| 17 | 6     | 0.30               | 0       |
| 18 | 7     | 0.20               | 0       |
| 19 | 8     | 0.15               | 0       |

- a) Từ bảng số liệu trên, áp dụng thuật toán Gradient Descent để viết hàm tính xác suất cho vay của một hồ sơ bất kỳ.
- b) Giả sử ngân hàng yêu cầu hồ sơ đạt 80% mới cho vay, hãy vẽ đường phân cách giữa hồ sơ cho vay và không cho vay. Từ đó xác định xem một người có mức lương là 9 triệu và kinh nghiệm làm việc là 0.5 năm thì có được vay hay không?

## Bài 4: Tương tự bài trên nhưng dùng **Accelerated Gradient Descent**.