# Yêu cầu:

1. Viết mã giả của thuật toán tìm phương trình hồi quy phi tuyến.

2. Đánh giá độ phức theo thời gian của thuật toán tìm phương trình hồi quy phi tuyến ở trên.

3. Với tập dữ liệu về Việt Nam tùy chọn, dùng MS Excel, ngôn ngữ R và Python, hãy tìm phương trình hồi quy phi tuyến.

# Thực hiện:

## Bảng phân công công việc:

|  |  |  |
| --- | --- | --- |
| **Thành viên** | | **Công việc phân công** |
| **Mã học viên** | **Họ và Tên** |
| 240104026 | Huỳnh Quốc Bữu | * Đánh giá độ phức theo thời gian của thuật toán tìm phương trình hồi quy phi tuyến * Tìm phương trình hồi quy phi tuyến bằng R * Tìm phương trình hồi quy phi tuyến bằng Python |
| 240104039 | Nguyễn Đức Linh | * Viết mã giả của thuật toán tìm phương trình hồi quy phi tuyến * Chọn tập dữ liệu về Việt Nam * Tiền xử lý dữ liệu * Tìm phương trình hồi quy phi tuyến bằng Excel |

## 1. Viết mã giả của thuật toán tìm phương trình hồi quy phi tuyến:

### 1.1 Mô tả bài toán:

Cho một tập dữ liệu gồm n quan sát:

Với:

* : là vector biến độc lập đầu vào có k phần tử
* : là biến phụ thuộc đầu ra

Bảng dữ liệu:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Mục tiêu: tìm hàm hồi qui phi tuyến sao cho:

Có nhiều mô hình hồi qui phi tuyến, trong bài tập nhóm chọn **mô hình hồi quy phi tuyến đa thức bậc d** (Polynomial d degree) [1]

Chọn là một hàm đa thức bậc d loại **polynomial d degree**  có dạng:

Với:

* : chỉ số mũ từng biến
* : tổng các bậc
* : là dạng của một đơn thức (nomial) tương tương ứng
* : hệ số tương ứng
* là tổng các nomial trong đa thức

Bảng mô tả các phần tử đa thức (nomial) ví dụ với k =3 và d = 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Tổng các nomial trong đa thức

Khi đó hàm hôì quy phi tuyến là:

Bảng mô tả các phần tử đa thức (nomial) tổng quát với k biến độc lập và bậc d

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Tổng các nomial trong đa thức:

Hàm hôì quy phi tuyến là:

Gọi ánh xạ

Phương trình hồi qui phi tuyến trên có thể viết lại như sau:

Do đó, bài toán:

với và tập dữ liệu

có thể biến đổi thành bài toán:

với và tập dữ liệu

Z được biến biến đổi qua ánh xạ

Dễ dàng nhận ra là hàm tuyến tính

Ta có thể giải bằng phương pháp bình phương tối thiểu (Ordinary Least Squares - OLS) [2]

Ta tính được:

### 1.2 Mã giả của thuật toán tìm phương trình hồi quy phi tuyến ở trên:

**Bước 1:** Nhập tập dữ liệu

với : là vector biến độc lập đầu vào có k phần tử

**Bước 2:** Xác định ma trận , vector

**Bước 3:** Chọn mô hình hồi quy phi tuyến là mô hình hồi qui đa thức bậc d

**Bước 4:** Tìm ma trận ma trận từ ma trận , trong đó

**Bước 5:** Tính ma trân chuyển vị

**Bước 6:** Tính ma trận

**Bước 7:** Tính ma nghịch đảo

**Bước 8:** Tính

**Bước 9:** Xuất kết quả tính được

## 2. Đánh giá độ phức theo thời gian của thuật toán tìm phương trình hồi quy phi tuyến ở trên

**Bước 1:** Nhập tập dữ liệu

với : là vector biến độc lập đầu vào có k phần tử

* BigO1 = O(1) điền thông tin vào

**Bước 2:** Xác định ma trận , vector

* BigO2 = O(xxx) điền thông tin vào

**Bước 3:** Chọn mô hình hồi quy phi tuyến là mô hình hồi qui đa thức bậc d

* BigO3 = O(1) điền thông tin vào

**Bước 4:** Tìm ma trận ma trận từ ma trận , trong đó

* BigO4 = O(xxx) điền thông tin vào

**Bước 5:** Tính ma trân chuyển vị

* BigO5 = O(xxx) điền thông tin vào

**Bước 6:** Tính ma trận

* BigO6 = O(xxx) điền thông tin vào

**Bước 7:** Tính ma nghịch đảo

* BigO7 = O(xxx) điền thông tin vào

**Bước 8:** Tính

* BigO8 = O(xxx) điền thông tin vào

**Bước 9:** Xuất kết quả tính được

* BigO9 = O(1) điền thông tin vào

**Độ phức tạp của tổng hợp của thuật toán tìm phương trình hồi qui phi tuyến là:**

BigO = O(xxx) điền thông tin vào

## 3. Với tập dữ liệu về Việt Nam tùy chọn, dùng MS Excel, ngôn ngữ R và Python, hãy tìm phương trình hồi quy phi tuyến

### 3.1 Chọn tập dữ liệu về Việt Nam

1. Chọn tập dữ liệu về giá nhà tại Việt Nam để phân tích
2. Nguồn dữ liệu: <https://www.kaggle.com/datasets/nguyentiennhan/vietnam-housing-dataset-2024/data> [3]
3. File dữ liệu: Github-Code-Team11/datasets/vietnam\_housing\_dataset.csv
4. Giới thiệu tập dữ liệu:

Bộ dữ liệu này chứa thông tin về các bất động sản nhà ở tại Việt Nam. Nó bao gồm các đặc điểm chi tiết của từng bất động sản như vị trí, đặc điểm vật lý, tình trạng pháp lý và nội thất, cùng với giá bán. Bộ dữ liệu này rất hữu ích cho việc phân tích bất động sản, xây dựng mô hình dự đoán giá và phân tích xu hướng thị trường.

Nguồn thu thập dữ liệu: Dữ liệu được thu thập (crawled) từ trang web batdongsan.vn.

1. Mô tả các cột dữ liệu:

* Address (Địa chỉ): Địa chỉ đầy đủ của bất động sản, bao gồm tên dự án (nếu có), đường, phường, quận và thành phố.
* Area (Diện tích): Tổng diện tích của bất động sản, đơn vị tính là mét vuông (m²).
* Frontage (Mặt tiền): Chiều rộng mặt trước của bất động sản, đơn vị tính là mét (m).
* Access Road (Đường vào): Chiều rộng của con đường dẫn vào bất động sản, đơn vị tính là mét (m).
* House Direction (Hướng nhà): Hướng chính mà mặt trước của căn nhà quay về (ví dụ: Đông, Tây, Nam, Bắc).
* Balcony Direction (Hướng ban công): Hướng mà ban công quay về.
* Floors (Số tầng): Tổng số tầng của căn nhà.
* Bedrooms (Số phòng ngủ): Số lượng phòng ngủ trong bất động sản.
* Bathrooms (Số phòng tắm): Số lượng phòng tắm trong bất động sản.
* Legal Status (Tình trạng pháp lý): Trạng thái pháp lý của bất động sản, ví dụ như có sổ hồng, đang trong hợp đồng mua bán, v.v.
* Furniture State (Tình trạng nội thất): Tình trạng trang bị nội thất, ví dụ như đầy đủ nội thất, nội thất cơ bản hoặc chưa có nội thất.
* Price (Giá bán): Giá của bất động sản, đơn vị tính là **tỷ đồng Việt Nam (VND)**.

### 3.2 Tiền xử lý dữ liệu:

điền thông tin vào

### 3.3 Tìm phương trình hồi quy phi tuyến bằng Excel:

điền thông tin vào

### 3.4 Tìm phương trình hồi quy phi tuyến bằng Python:

### 3.5 Tìm phương trình hồi quy phi tuyến bằng R:

## Kiểm tra kết quả với ChatGPT

điền thông tin vào

# References

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

|  |  |
| --- | --- |
| [[1]]([1] G. James, D. Witten, T. Hastie và R. Tibshirani, An Introduction to Statistical Learning with Applications in R, New York: Springer, 2021. [2] D. C. P. E. A. &. V. G. G. Montgomery, Introduction to Linear Regression Analysis, 5th Edition, Wiley, 2012. [3] N. T. Nhan, \“Kaggle,\” [Trực tuyến]. Available: https://www.kaggle.com/datasets/nguyentiennhan/vietnam-housing-dataset-2024/data.   ) | [G. James, D. Witten, T. Hastie và R. Tibshirani, An Introduction to Statistical Learning with Applications in R, New York: Springer, 2021.]([1] G. James, D. Witten, T. Hastie và R. Tibshirani, An Introduction to Statistical Learning with Applications in R, New York: Springer, 2021. [2] D. C. P. E. A. &. V. G. G. Montgomery, Introduction to Linear Regression Analysis, 5th Edition, Wiley, 2012. [3] N. T. Nhan, \“Kaggle,\” [Trực tuyến]. Available: https://www.kaggle.com/datasets/nguyentiennhan/vietnam-housing-dataset-2024/data.   ) |
| [[2]]([1] G. James, D. Witten, T. Hastie và R. Tibshirani, An Introduction to Statistical Learning with Applications in R, New York: Springer, 2021. [2] D. C. P. E. A. &. V. G. G. Montgomery, Introduction to Linear Regression Analysis, 5th Edition, Wiley, 2012. [3] N. T. Nhan, \“Kaggle,\” [Trực tuyến]. Available: https://www.kaggle.com/datasets/nguyentiennhan/vietnam-housing-dataset-2024/data.   ) | [D. C. P. E. A. &. V. G. G. Montgomery, Introduction to Linear Regression Analysis, 5th Edition, Wiley, 2012.]([1] G. James, D. Witten, T. Hastie và R. Tibshirani, An Introduction to Statistical Learning with Applications in R, New York: Springer, 2021. [2] D. C. P. E. A. &. V. G. G. Montgomery, Introduction to Linear Regression Analysis, 5th Edition, Wiley, 2012. [3] N. T. Nhan, \“Kaggle,\” [Trực tuyến]. Available: https://www.kaggle.com/datasets/nguyentiennhan/vietnam-housing-dataset-2024/data.   ) |
| [[3]]([1] G. James, D. Witten, T. Hastie và R. Tibshirani, An Introduction to Statistical Learning with Applications in R, New York: Springer, 2021. [2] D. C. P. E. A. &. V. G. G. Montgomery, Introduction to Linear Regression Analysis, 5th Edition, Wiley, 2012. [3] N. T. Nhan, \“Kaggle,\” [Trực tuyến]. Available: https://www.kaggle.com/datasets/nguyentiennhan/vietnam-housing-dataset-2024/data.   ) | [N. T. Nhan, “Kaggle,” [Trực tuyến]. Available: https://www.kaggle.com/datasets/nguyentiennhan/vietnam-housing-dataset-2024/data.]([1] G. James, D. Witten, T. Hastie và R. Tibshirani, An Introduction to Statistical Learning with Applications in R, New York: Springer, 2021. [2] D. C. P. E. A. &. V. G. G. Montgomery, Introduction to Linear Regression Analysis, 5th Edition, Wiley, 2012. [3] N. T. Nhan, \“Kaggle,\” [Trực tuyến]. Available: https://www.kaggle.com/datasets/nguyentiennhan/vietnam-housing-dataset-2024/data.   ) |