

ĐỒ ÁN 1 - HỒI QUY TUYẾN TÍNH

1 Về tập dữ liệu

Cho tập dữ liệu về một trang mua bán xe hơi cũ. Tập dữ liệu sẽ gồm 2 phần: phần X sẽ gồm các thông tin về xe, phần Y là giá mua bán xe. Dữ liệu trong phần X được mô tả như sau:

- id: Mã số định danh
- manufacturer: Tên nhà sản xuất xe hơi
- model: Tên mẫu xe
- transmission: Loại hộp số
- color: Màu xe
- odometer: Số km mà xe đã đi
- year: Năm sản xuất
- engineFuel: Loại nhiên liệu động cơ
- engineType: Loại động cơ
- engineCapacity: Dung tích động cơ tính bằng lít
- bodyType: Dạng thân xe
- drivetrain: Loại hệ thống dẫn động
- photos: Số lượng hình ảnh được chụp về xe
- feature0, ..., feature9: Các đặc trưng khác của xe

Dữ liệu trong phần Y được mô tả như sau:

- id: Mã số định danh
- price: Giá xe tính theo USD.

2 Về đề bài

- Sinh viên áp dụng các kiến thức về hồi quy tuyến tính để huấn luyện mô hình dựa vào tập dữ liệu ở trên và đưa ra dự đoán về giá xe.
- Việc chọn mô hình như thế nào cho phù hợp là do sinh viên tự quyết định. Tuy nhiên, sinh viên cần phải giải thích lý do tại sao bản thân lại chọn mô hình như vậy.
- Trên Moodle, sinh viên sẽ được cung cấp tập dữ liệu training set để huấn luyện mô hình. Khi chấm vấn đáp, sinh viên sẽ được cung cấp tập dữ liệu test set để đánh giá độ tốt của mô hình.
- Ngoài những yêu cầu trên, sinh viên NÊN chuẩn bị thêm các chương trình dạng dòng lệnh như sau:
 - + Để huấn luyện, sinh viên truyền vào đường dẫn phần X và Y của tập dữ liệu và đường dẫn mô hình thu được sau khi huấn luyện. Ví dụ: `train.exe X.csv Y.csv output.model`
 - + Để đánh giá, sinh viên truyền vào đường dẫn mô hình đã được huấn luyện, đường dẫn phần X của tập dữ liệu và đường dẫn phần Y mà mô hình sẽ dự đoán. Ví dụ: `test.exe test.model X.csv predictY.csv`

3 Về quy định nộp bài

Sinh viên cần nộp đầy đủ các thành phần sau:

- File báo cáo đồ án: report.pdf
- Thư mục chứa mã nguồn chương trình: source
- Trong thư mục source, cần có file README.TXT hướng dẫn chi tiết cách chạy các chức năng huấn luyện mô hình và kiểm tra mô hình.