Bài tập chương I

Thành viên nhóm:

Đỗ Lý Hoàng Nam

Nguyễn Anh Tú

Hoàng Thế Phong

1. Chuẩn bị dữ liệu

Mô tả:

Bộ dữ liệu Boston Housing là một tập dữ liệu phổ biếm. Nó dự đoán giá nhà trung bình (MEDV) dựa trên 13 đặc trưng (features) khác nhau, như tỷ lệ tội phạm, chất lượng không khí, số phòng, v.v.

Thao tác:

Nạp dữ liệu Boston Housing từ thư viện scikit-learn.

Chia dữ liệu thành:

X: các đặc trưng (features).

y: giá trị mục tiêu (target) - MEDV (Median value of owner-occupied homes).

Chia dữ liệu thành tập huấn luyện và tập kiểm tra theo tỷ lệ (thường là 80:20).

2. Cài đặt mô hình KNN cho bài toán hồi quy

Mô tả:

KNN hồi quy tính giá trị mục tiêu của một điểm dữ liệu mới bằng cách lấy trung bình giá trị mục tiêu của K điểm lân cận nhất.

Thao tác:

Import KNeighborsRegressor từ thư viện scikit-learn.

Xác định giá trị K (ví dụ: K=5).

Khởi tạo mô hình hồi quy KNN với giá trị K.

3. Huấn luyện và dự đoán

Mô tả:

Dữ liệu huấn luyện sẽ được mô hình lưu trữ để tính khoảng cách khi dự đoán.

Mô hình tính toán giá trị dự đoán bằng cách lấy trung bình các giá trị mục tiêu của K hàng xóm gần nhất.

Thao tác:

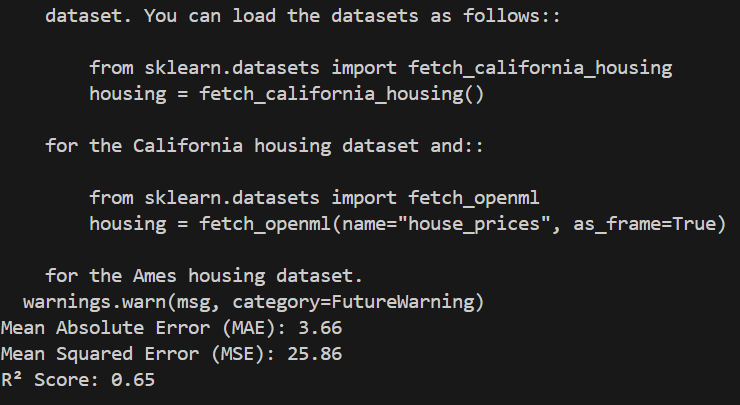
Huấn luyện mô hình trên tập dữ liệu huấn luyện (fit()).

Sử dụng mô hình để dự đoán giá trị mục tiêu trên tập kiểm tra (predict()).

4. Đánh giá mô hình

Mô tả:

Sử dụng các thước đo phổ biến cho bài toán hồi quy để đánh giá độ chính xác của mô hình



5. Hiển thị kết quả

Mô tả:

Trực quan hóa kết quả dự đoán so với giá trị thực tế trên tập kiểm tra để dễ dàng đánh giá hiệu suất của mô hình.

Thao tác:

Vẽ biểu đồ so sánh giá trị thực tế (y\_test) và giá trị dự đoán (y\_pred) bằng matplotlib.

Vẽ đồ thị biểu diễn MAE hoặc R² khi thay đổi giá trị K.

