1.Cho các điểm dữ liệu (x,y) = {(1,1), (2,3), (3, 2), (4,4), (5, 4)}.

a) Tính MSE với model y = 2x+1.
b) Cho nhập a, b tính MSE với model y = ax+b.
c) Dùng thư viện sklearn tìm model linear regerssion phù hợp nhất, in ra các hệ số coefficient, intercep của model. Tính MSE, score (https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html#sklearn.linear_model.LinearRegression)

2.Load file dữ liệu data_linear.csv trong mục data:

a) Visualize dữ liệu dạng scatter.
b) Dùng sklearn tìm model linear regression dự đoán giá nhà theo diện tích, in ra các hệ số, tính MSE của model.
c) Dự đoán các căn có diện tích [50, 120, 150] có giá bao nhiêu.
d) Xem việc dự đoán của model có tốt hay không, vẽ đường thẳng model trên dữ liệu scatter ở trên.

3.Load dữ liệu file data_square.csv trong mục data:

a) Visualize dữ liệu dạng scatter
b) Dùng sklearn tìm model dự đoán giá nhà theo diện tích, in ra các hệ số, tính MSE của model.
c) Model linear regression ở phần b có tốt không? Tại sao?

Dùng dữ liệu bài 3, mình biết là model cần thiết là parabol y = ax^2 + bx + c chứ linear model không đủ tốt để mô tả dữ liệu.

a) Khi load X mình sẽ tạo thêm 1 cột X^2 cho dữ liệu, giờ thành [X, X^2] cho X, Y giữ nguyên. Dùng linear regression để fit vào model.
b) Visualize xem prediction của model có tốt hơn hay không (vẽ hàm parabol)?