

Chương 2 - Phần 2

TS Phạm Tuấn

Bài tập nhóm số 2

1. Download file dữ liệu tại [đây](#). Dùng Pandas đọc file.
2. Đọc dữ liệu từ file đó và chuyển 2 cột **Height** và **Weight** từ **inch** và **pound** về **cm** và **kg**. In ra Head()
3. Vẽ dữ liệu lên trục toạ độ với **2 màu sắc khác nhau biểu thị cho 2 giới tính**.
4. Tạo mô hình Hồi quy tuyến tính cho tập dữ liệu trên dùng Sklearn. Với đầu vào là chiều cao, dữ liệu nhãn là cân nặng.
5. **Tìm toạ độ nhiễu, in ra và loại bỏ nhiễu.**
6. Chia dữ liệu thành **2 phần** train_set (80%) và test_set (20%) chia đều cho cả 2 giới tính. (in ra số lượng dữ liệu cho nam và nữ ở 2 tập dữ liệu trên)
7. Tạo mô hình Hồi quy tuyến tính **cho tập train_set** và tính R2_score (dùng sklearn).
8. Tính điểm R2_score **cho tập test_set**.
9. Viết Hàm để nhập chiều cao, cân nặng sau đó dự đoán tình trạng người đó như thế nào, dựa trên sai số, nếu sai số càng thấp thì cơ thể càng cân đối.