Họ và tên: Nguyễn Nhật Trường

MSSV: 20522087

Bài tập lí thuyết 01: Chuẩn hóa giá trị liên tục

# Có những phương pháp chuẩn hóa giá trị liên tục nào trong Sklearn? Công thức từng phương pháp chuẩn hóa là gì? Khi nào ta nên sử dụng hàm chuẩn hóa nào?

|  |  |  |
| --- | --- | --- |
| Phương pháp chuẩn hóa | Công thức | Khi nào nên sử dụng |
| MaxAbsScaler |  | Sử dụng hàm chuẩn hóa MaxAbsScaler khi độ lệnh chuẩn của các đặc trưng quá nhỏ và muốn giữ lại giá trị 0 trong ma trận thưa thớt. |
| MinMaxScaler |  | Khi muốn giữ nguyên hình dạng phân phối ban đầu. Không làm giảm tầm quan trọng của các yếu tố ngoại lai. Ít làm gián đoạn thông tin của dữ liệu gốc. |
| Normalizer | Mỗi hàng được chuẩn hóa bằng cách áp dụng chuẩn hóa l2 (Euclidian). Nếu mỗi phần tử được bình phương và tính tổng thì tổng sẽ bằng 1. Cũng có thể dùng chuẩn hóa l1 (Mahattan).  Với z có thể là:   * : giá trị trên từng dòng | Dùng khi muốn chuẩn hóa mỗi dòng, không dùng để chuẩn hóa các cột. Hiếm khi được sử dụng. |
| RobustScaler | Với:   * giá trị trên từng cột * Median: giá trị trung vị | Sử dụng RobustScaler khi dữ liệu có các ngoại lai (outliers) và mong muốn làm giảm tầm ảnh hưởng của các outliers với dữ liệu của mình. |
| StandardScaler | Với:   * giá trị trung bình * : độ lệch chuẩn * : giá trị trên từng cột | Sử dụng StandardScaler khi mong muốn mỗi đặc trưng có giá trị trung bình bằng 0 và độ lệch chuẩn bằng 1. Nếu muốn dữ liệu được phân phối ổn định với việc chuyển đổi dữ liệu. |

Tài liệu tham khảo:

[Min Max Normalization in data mining | T4Tutorials.com](https://t4tutorials.com/min-max-normalization-of-data-in-data-mining/)

[API Reference — scikit-learn 1.1.2 documentation](https://scikit-learn.org/stable/modules/classes.html#module-sklearn.preprocessing)