

Các phương pháp chuẩn hoá:

1. Minmax Scaler

- Là phương pháp đơn giản nhất để đưa các Feature về cùng khoảng $[-1;1]$ hay $[0;1]$
- Công thức tính $x' = \frac{x - x_{min}}{x_{max} - x_{min}}$

Trong đó:

- x' : giá trị sau chuẩn hoá.
- x : giá trị ban đầu.
- Sử dụng khi dữ liệu xác định được miền giá trị

2. Standard Scaler

- Biến đổi dữ liệu sao cho phân phối của kết quả có giá trị trung bình (μ) bằng 0 và độ lệch chuẩn (σ) bằng 1
- Công thức tính: $x' = \frac{x - x_{mean}}{\sigma}$

Trong đó:

- x' : giá trị sau chuẩn hoá.
- x : giá trị ban đầu.
- x_{mean} : giá trị mean của bộ dữ liệu.
- σ : phương sai (standard deviation).
- Sử dụng khi dữ liệu đầu vào có phân phối chuẩn

3. Normalize

- Công thức tính $x' = \frac{x - x_{min}}{x_{max} - x_{min}}$

Trong đó:

- x' : giá trị sau chuẩn hoá.
- x : giá trị ban đầu.
- x_{min} : giá trị min của bộ dữ liệu.
- Trong phương pháp này, giá trị lớn nhất có được là 1 và nhỏ nhất là 0. Nên dữ liệu nằm trong đoạn $[0;1]$.