

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



BÀI TẬP MÔN LẬP TRÌNH PYTHON CHO MÁY HỌC

KHOA: KHOA HỌC MÁY TÍNH

HOMEWORK: CHUẨN HOÁ GIÁ TRỊ LIÊN TỤC

Giảng viên: Nguyễn Vinh Tiệp

Nhóm thực hiện:

- 1. Trương Thành Thắng – 20521907**

Đề: 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?

Bài làm:

Phương pháp		Công thức	Trường hợp dùng
MinMaxScaler		$x'_{:,c} = \frac{x_{:,c} - \min\{x_{:,c}\}}{\max\{x_{:,c}\} - \min\{x_{:,c}\}} * (fr.max - fr.min) + fr.min$	Khi muốn giảm phạm vi phân bố của dữ liệu mà vẫn giữ nguyên hình dạng phân phối ban đầu.
MaxAbsScaler		$x'_{:,c} = \frac{x_{:,c}}{\max\{x_{:,c}\}}$	Khi độ lệch chuẩn của các đặc trưng quá nhỏ.
StandardScaler		$x'_{:,c} = \frac{x_{:,c} - \mu_{:,c}}{\sigma_{:,c}}$	<p>Thông tin dữ liệu có ý nghĩa nằm trong mối quan hệ giữa các đặc trưng từ mẫu này sang mẫu khác.</p> <p>Dữ liệu có giá trị âm.</p> <p>Dùng trong bài toán Phân lớp tốt hơn Hồi quy.</p>
RobustScaler		$x'_{:,c} = \frac{x_{:,c} - median\{x_{:,c}\}}{x_{:,c_{75}} - x_{:,c_{25}}}$	Loại bỏ các outlier.
Normalizer	L1	$x'_{r,:} = \frac{x_{r,:}}{\sum_c x_{r,c} }$	Thông tin dữ liệu có ý nghĩa nằm trong mối quan hệ giữa các đặc trưng từ đặc trưng này sang đặc trưng khác.
	L2	$x'_{r,:} = \frac{x_{r,:}}{\sqrt{\sum_c x_{r,c} ^2}}$	
	Max	$x'_{r,:} = \frac{x_{r,:}}{\max\{x_{r,:}\}}$	Dùng trong bài toán Hồi quy tốt hơn trong Phân lớp.

PowerTransformer	Box-Cox	$x'_{:,c} = \begin{cases} \frac{x_{:,c} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(x_{:,c}) & \text{if } \lambda = 0 \end{cases}$	<p>Làm cho dữ liệu giống với Gaussian hơn nhằm ổn định phương sai và giảm thiểu độ lệch.</p> <p>Box-Cox chỉ áp dụng cho dữ liệu dương</p>
	Yeo-Johnson	$x'_{:,c} = \begin{cases} \frac{(x_{:,c} + 1)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, x_{:,c} \geq 0 \\ \ln(x_{:,c} + 1) & \text{if } \lambda = 0, x_{:,c} \geq 0 \\ \frac{(-x_{:,c} + 1)^{x-\lambda} - 1}{\lambda - 2} & \text{if } \lambda \neq 2, x_{:,c} < 0 \\ -\ln(-x_{:,c} + 1) & \text{if } \lambda = 2, x_{:,c} < 0 \end{cases}$	