

Feature Scaling



Feature Scaling

Data Transformation

Normalization & Standardization



Agenda

- Why Data Scaling / Transformation?
- Transformation Techniques
- Normalization
- Standardization
- Related Topics

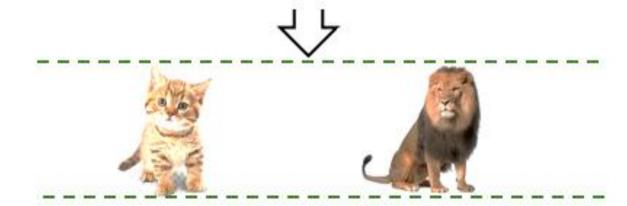
.







FEATURE SCALING





Scaling / Transformation

Age of Customer (Years)	Purchase (Latest Car model)
32	No
26	No
29	Yes
37 .	No
42	No
47	Yes
30	No
39	No
27	No
	32 26 29 37 . 42 47 30 39



Scaling / Transformation

Monthly Income (\$)	Age of Customer (Years)	Purchase (Latest Car model)
1800 18	32	No
700 - 7-	26	No
2750 27.5	29	Yes
3300 33	37	No
6000 60-	42	No
4500 -45	47	Yes
920 9.2	30	No
5050 50.5	39	No
1400 14	27	No
0-1	0-1	



MIN MAX SCALER (Normalization)

1 27 48000 2 30 54000 3 38 61000 4 40 70000 5 35 58000 6 31 52000 7 48 79000 8 50 83000 0 . 0. 0 . 13043478 0 .17142857 0 .47826087 0 .37142857 0 .56521739 0 .628571439 0 .34782609 0 .28571429 0 .17391304 0 .11428571 0 .91304348 0 .88571429 1 . 1 .		Age	Salary	Age	e Salary
2 30 54000 3 38 61000 4 40 70000 5 35 58000 6 31 52000 7 48 79000 8 50 83000 0 .13043478 0 .17142857 0 .47826087 0 .37142857 0 .56521739 0 .62857143 0 .34782609 0 .28571429 0 .17391304 0 .11428571 0 .91304348 0 .88571429 1 .	0	44	72000	0.7391	3043 0.68571429
3 38 61000 4 40 70000 5 35 58000 6 31 52000 7 48 79000 8 50 83000 0 .47826087 0 .37142857 0 .56521739 0 .628571439 0 .34782609 0 .28571429 0 .17391304 0 .11428571 0 .91304348 0 .88571429 1 . 1 .	1	27	48000	0.	0.
4 40 70000 0.56521739 0.62857143 5 35 58000 0.34782609 0.28571429 6 31 52000 0.17391304 0.11428571 7 48 79000 0.91304348 0.88571429 8 50 83000 1. 1.	2	30	54000	0.1304	13478 0.17142857
5 35 58000 6 31 52000 7 48 79000 8 50 83000 Scaling 0.34782609 0.28571429 0.17391304 0.11428571 0.91304348 0.88571429 1. 1.	3	38	61000	0.4782	26087 0.37142857
Scaling 0.17391304 0.11428571 0.91304348 0.88571429 1. 1.	4	40	70000	0.5652	21739 0.62857143
7 48 79000 0.91304348 0.88571429 8 50 83000 1. 1.	5	35	58000	· · · · · · · · · · · · · · · · · · ·	82609 0.28571429
8 50 83000 1. 1.	6	31	52000	Scaling 0.1739	91304 0.11428571
'a 42470064	7	48	79000	0.9130	04348 0.88571429
9 37 67000 [0.43478261 0.54285714	8	50	83000	1.	1.
	9	37	67000	0.434	78261 0.54285714



Scaling / Transformation Techniques

- Normalization
- Standardization
- Quantile transformation
- Log transformation
- Winsorization
- Power transformation
- Unit Vector Scaling



Normalization

Transformation

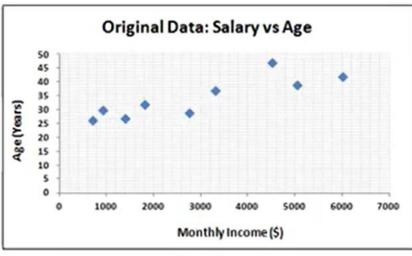
Mathematical Formula

$$X_{\text{transformed}} = \underbrace{X - X_{\text{min}}}_{X_{\text{max}} - X_{\text{min}}}$$

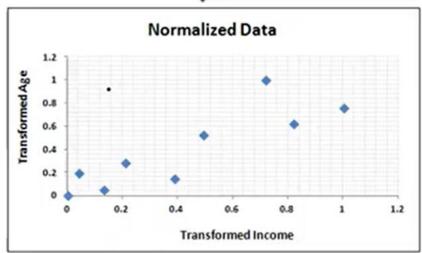
Python Function

Sklearn.preprocessing.

MinMaxScaler()









Transformation

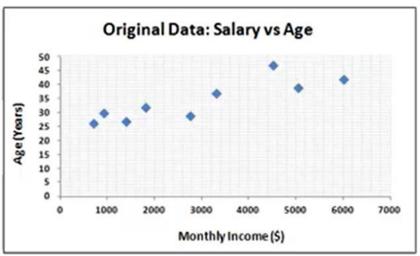
Mathematical Formula

$$X_{transformed} = X - \mu$$

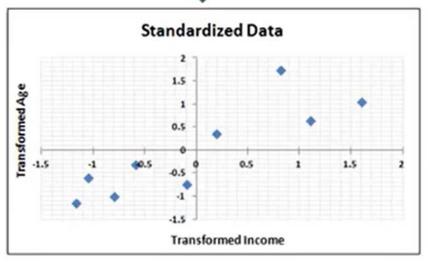
$$X_{transformed} = \frac{X_i - X_{mean}}{S_{tandard Deviation}}$$

Python Function

Sklearn.preprocessing. StandardScaler()









Feature Scaling

Examples of Algorithms where Feature Scaling matters:

- 1. K-Means uses the Euclidean distance measure here feature scaling matters.
- 2. K-Nearest-Neighbours also require feature scaling.
- 3. **Principal Component Analysis (PCA)**: Tries to get the feature with maximum variance, here too feature scaling is required.
- 4. Gradient Descent: Calculation speed increase as Theta calculation becomes faster after feature scaling.

Note: Naive Bayes, Linear Discriminant Analysis, and Tree-Based models are not affected by feature scaling.



Related Topics

- One-hot Encoding
- Missing value treatment
- Outlier treatment
- Feature Engineering
- Dimensionality Reduction