Feature Scaling

Aksadur Rahman

aksadur@yahoo.com

DATA SCIENCE

Feature Scaling

Data Transformation

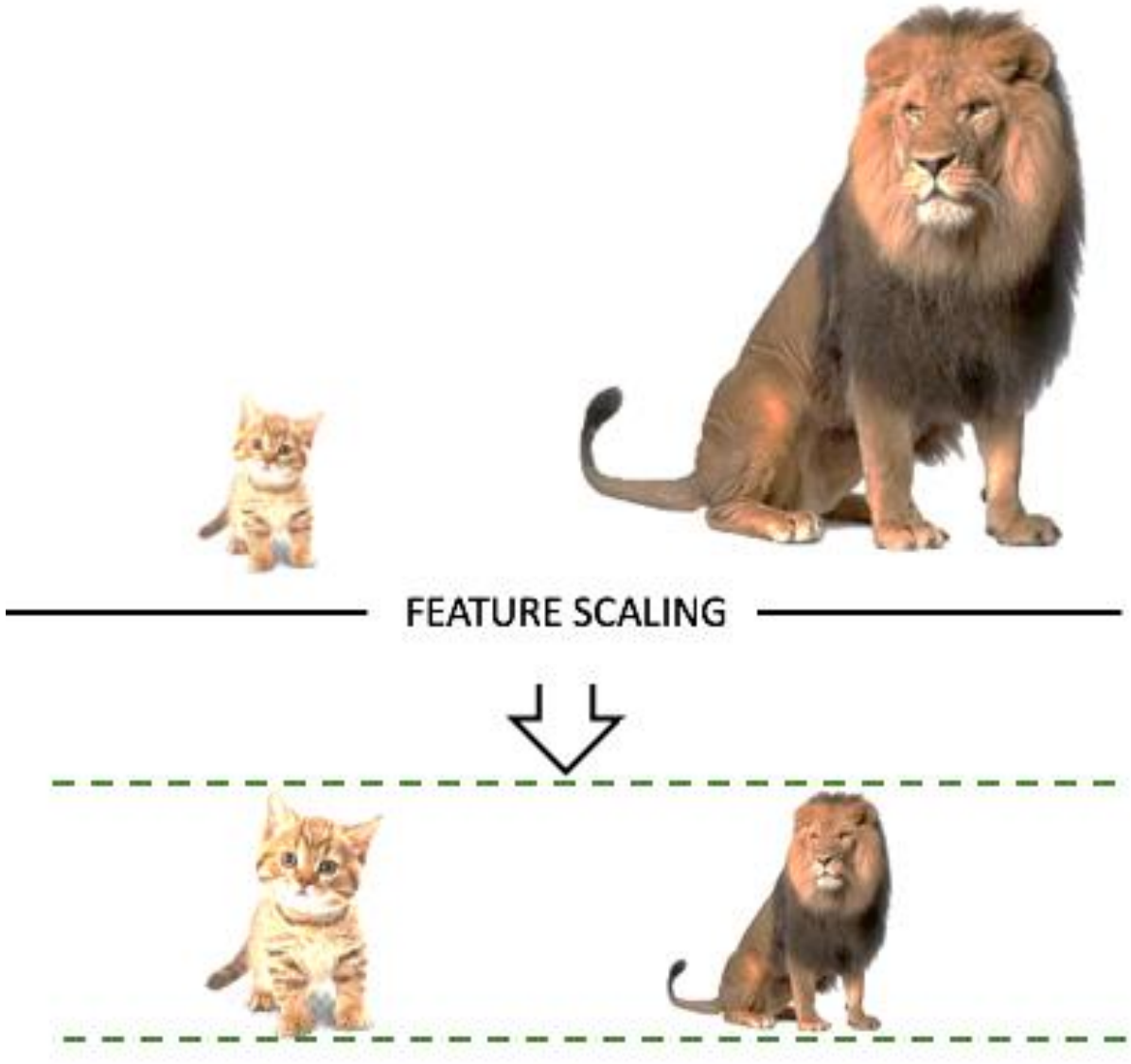
Normalization & Standardization

DATA SCIENCE

Agenda

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

DATA SCIENCE



DATA SCIENCE

Scaling / Transformation

Monthly Income ($)	Age of Customer (Years)	Purchase (Latest Car model)
1800	32	No
700	26	No
2750	29	Yes
3300	37	No
6000	42	No
4500	47	Yes
920	30	No
5050	39	No
1400	27	No

DATA SCIENCE

Scaling / Transformation

Monthly Income ($)	Age of Customer (Years)	Purchase (Latest Car model)
1800	32	No
700	26	No
2750	29	Yes
3300	37	No
6000	42	No
4500	47	Yes
920	30	No
5050	39	No
1400	27	No

DATA SCIENCE

MIN MAX SCALER (Normalization)

	Age	Salary
0	44	72000
1	27	48000
2	30	54000
3	38	61000
4	40	70000
5	35	58000
6	31	52000
7	48	79000
8	50	83000
9	37	67000

Scaling

Age	Salary
0.73913043	0.68571429
0.	0.
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.	1.
0.43478261	0.54285714

DATA SCIENCE

Scaling / Transformation Techniques

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

DATA SCIENCE

Normalization

Transformation

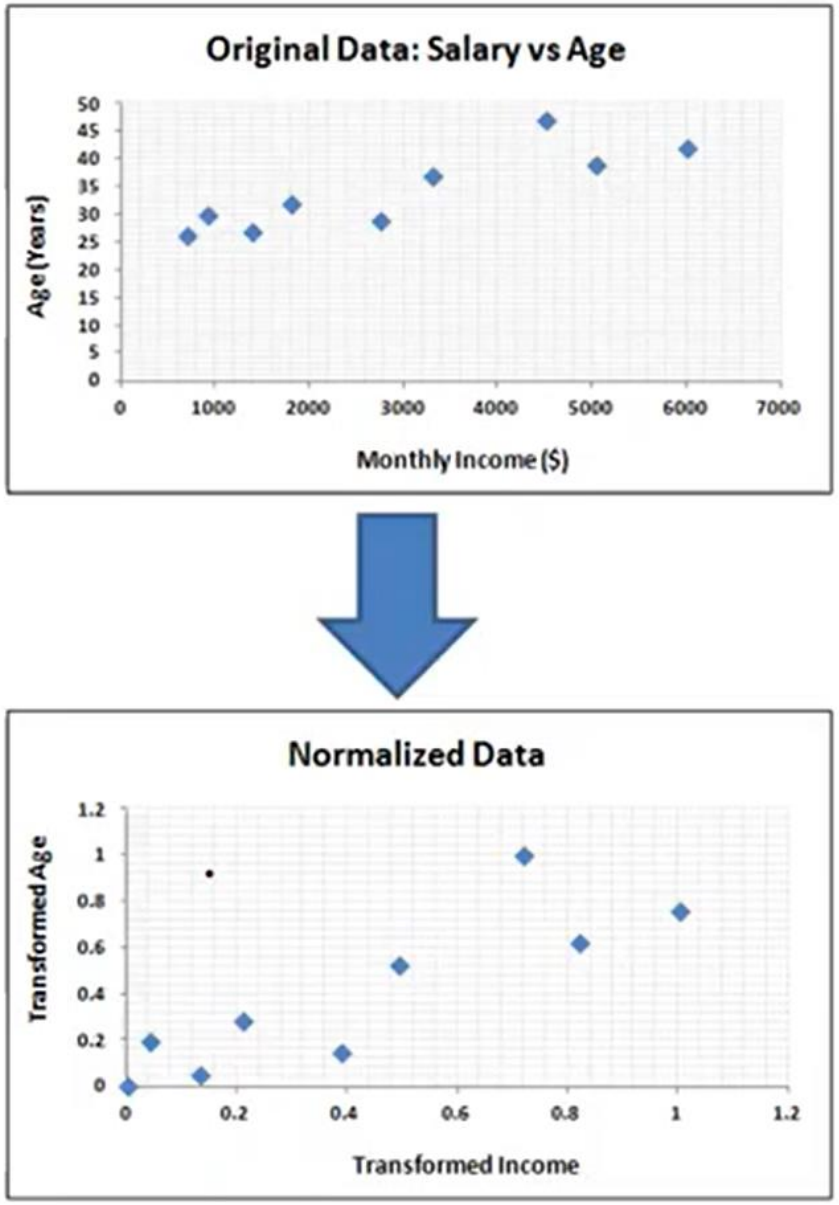
Mathematical Formula

$$X_{transformed} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Python Function

Sklearn.preprocessing.

MinMaxScaler()



DATA SCIENCE

Standardization

Transformation

Mathematical Formula

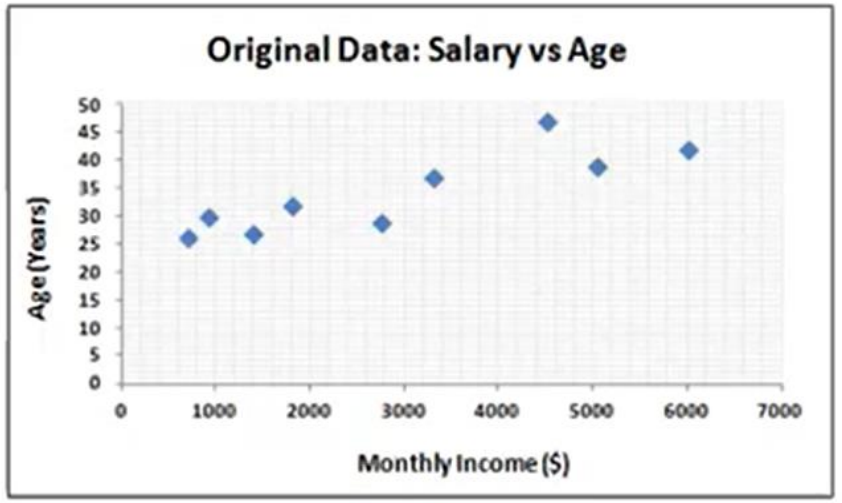
- $X_{transformed} = \frac{X - \mu}{\sigma}$

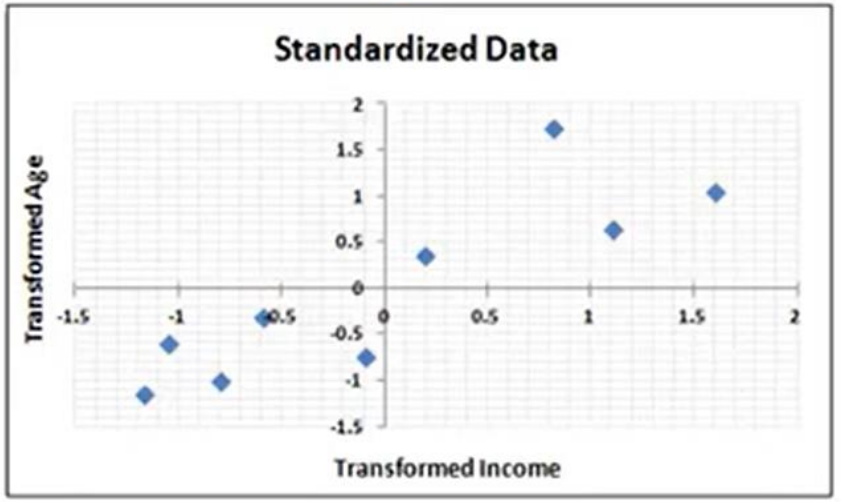
$$X_{transformed} = \frac{X_i - X_{mean}}{\text{Standard Deviation}}$$

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.

DATA SCIENCE

Related Topics

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