Feature Scaling in Machine Learning

Bindeshwar Singh Kushwaha
Postnetwork Academy

Feature Scaling

- Machine Learning algorithms often struggle when input numerical features have vastly different scales.

Feature Scaling

- Machine Learning algorithms often struggle when input numerical features have vastly different scales.
- Example: total number of rooms ranges from 6 to 39,320 while median incomes range from 0 to 15.

Feature Scaling

- Machine Learning algorithms often struggle when input numerical features have vastly different scales.
- Example: total number of rooms ranges from 6 to 39,320 while median incomes range from 0 to 15.
- Two common methods to scale features:

Feature Scaling

- Machine Learning algorithms often struggle when input numerical features have vastly different scales.
- Example: total number of rooms ranges from 6 to 39,320 while median incomes range from 0 to 15.
- Two common methods to scale features:
  - Min-Max Scaling (Normalization)

Feature Scaling

- Machine Learning algorithms often struggle when input numerical features have vastly different scales.
- Example: total number of rooms ranges from 6 to 39,320 while median incomes range from 0 to 15.
- Two common methods to scale features:
  - Min-Max Scaling (Normalization)
  - Standardization

Min-Max Scaling (Normalization)

- Rescales values to a fixed range, typically $[0, 1]$.

Min-Max Scaling (Normalization)

- Rescales values to a fixed range, typically $[0, 1]$.
- Formula: $x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$

Min-Max Scaling (Normalization)

- Rescales values to a fixed range, typically [0, 1].
- Formula: $x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$
- Implemented in Scikit-Learn via `MinMaxScaler`.

Min-Max Scaling (Normalization)

- Rescales values to a fixed range, typically [0, 1].
- Formula: $x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$
- Implemented in Scikit-Learn via `MinMaxScaler`.
- Use `feature_range` to change the output range if needed.

Standardization

- Subtract the mean and divide by the standard deviation.

Standardization

- Subtract the mean and divide by the standard deviation.
- Produces features with zero mean and unit variance.

Standardization

- Subtract the mean and divide by the standard deviation.
- Produces features with zero mean and unit variance.
- Formula: $x_{\text{standard}} = \frac{x-\mu}{\sigma}$

Standardization

- Subtract the mean and divide by the standard deviation.
- Produces features with zero mean and unit variance.
- Formula: $x_{\text{standard}} = \frac{x-\mu}{\sigma}$
- Doesn't bound values to a fixed range.

Standardization

- Subtract the mean and divide by the standard deviation.
- Produces features with zero mean and unit variance.
- Formula: $x_{\text{standard}} = \frac{x-\mu}{\sigma}$
- Doesn't bound values to a fixed range.
- Less affected by outliers compared to min-max scaling.

Standardization

- Subtract the mean and divide by the standard deviation.
- Produces features with zero mean and unit variance.
- Formula: $x_{\text{standard}} = \frac{x-\mu}{\sigma}$
- Doesn't bound values to a fixed range.
- Less affected by outliers compared to min-max scaling.
- Scikit-Learn uses `StandardScaler` for this.

Comparison Example

- A district had a median income equal to 100 (error).

Comparison Example

- A district had a median income equal to 100 (error).
- **Min-Max Scaling:** All values shift from [0–15] to [0–0.15], affected by the outlier.

Comparison Example

- A district had a median income equal to 100 (error).
- **Min-Max Scaling:** All values shift from [0–15] to [0–0.15], affected by the outlier.
- **Standardization:** Less influenced by the outlier due to use of mean and standard deviation.

Impact of Scaling on Sample Data

Feature Vector	Rooms	Income
Original	1000	4.5
Original	2000	6.0
Original	3000	8.5
Min-Max Scaled	0.00	0.00
Min-Max Scaled	0.50	0.43
Min-Max Scaled	1.00	1.00
Standardized	-1.22	-1.13
Standardized	0.00	-0.23
Standardized	1.22	1.36

Impact of Scaling on Sample Data

Feature Vector	Rooms	Income
Original	1000	4.5
Original	2000	6.0
Original	3000	8.5
Min-Max Scaled	0.00	0.00
Min-Max Scaled	0.50	0.43
Min-Max Scaled	1.00	1.00
Standardized	-1.22	-1.13
Standardized	0.00	-0.23
Standardized	1.22	1.36

- Min-Max bounds values between 0 and 1, but is affected by scale difference.

Impact of Scaling on Sample Data

Feature Vector	Rooms	Income
Original	1000	4.5
Original	2000	6.0
Original	3000	8.5
Min-Max Scaled	0.00	0.00
Min-Max Scaled	0.50	0.43
Min-Max Scaled	1.00	1.00
Standardized	-1.22	-1.13
Standardized	0.00	-0.23
Standardized	1.22	1.36

- Min-Max bounds values between 0 and 1, but is affected by scale difference.
- Standardization centers around 0, maintains relative distance and handles variance better.

Python Code Example: Feature Scaling with Scikit-Learn

```
from sklearn.preprocessing import MinMaxScaler, StandardScaler
import numpy as np

# Sample dataset
X = np.array([[1000, 4.5],
[2000, 6.0],
[3000, 8.5]])

# Min-Max Scaling
minmax_scaler = MinMaxScaler()
X_minmax = minmax_scaler.fit_transform(X)

# Standardization
standard_scaler = StandardScaler()
X_standard = standard_scaler.fit_transform(X)

print("Min-Max Scaled:\n", X_minmax)
print("Standardized:\n", X_standard)
```

Python Code Example: Feature Scaling with Scikit-Learn

```
from sklearn.preprocessing import MinMaxScaler, StandardScaler
import numpy as np

# Sample dataset
X = np.array([[1000, 4.5],
[2000, 6.0],
[3000, 8.5]])

# Min-Max Scaling
minmax_scaler = MinMaxScaler()
X_minmax = minmax_scaler.fit_transform(X)

# Standardization
standard_scaler = StandardScaler()
X_standard = standard_scaler.fit_transform(X)

print("Min-Max Scaled:\n", X_minmax)
print("Standardized:\n", X_standard)
```

- `MinMaxScaler` scales features to $[0, 1]$ range.

Python Code Example: Feature Scaling with Scikit-Learn

```
from sklearn.preprocessing import MinMaxScaler, StandardScaler
import numpy as np

# Sample dataset
X = np.array([[1000, 4.5],
[2000, 6.0],
[3000, 8.5]])

# Min-Max Scaling
minmax_scaler = MinMaxScaler()
X_minmax = minmax_scaler.fit_transform(X)

# Standardization
standard_scaler = StandardScaler()
X_standard = standard_scaler.fit_transform(X)

print("Min-Max Scaled:\n", X_minmax)
print("Standardized:\n", X_standard)
```

- `MinMaxScaler` scales features to [0, 1] range.
- `StandardScaler` centers features around 0 with unit variance.

Reach PostNetwork Academy

Website

www.postnetwork.co

Reach PostNetwork Academy

Website

www.postnetwork.co

YouTube Channel

www.youtube.com/@postnetworkacademy

Reach PostNetwork Academy

Website

www.postnetwork.co

YouTube Channel

www.youtube.com/@postnetworkacademy

Facebook Page

www.facebook.com/postnetworkacademy

Reach PostNetwork Academy

Website

www.postnetwork.co

YouTube Channel

www.youtube.com/@postnetworkacademy

Facebook Page

www.facebook.com/postnetworkacademy

LinkedIn Page

www.linkedin.com/company/postnetworkacademy

Reach PostNetwork Academy

Website

www.postnetwork.co

YouTube Channel

www.youtube.com/@postnetworkacademy

Facebook Page

www.facebook.com/postnetworkacademy

LinkedIn Page

www.linkedin.com/company/postnetworkacademy

GitHub Repositories

www.github.com/postnetworkacademy

Thank You!