Unsupervised Machine Learning

Unsupervised Machine Learning

What's the difference?

- No labels / ground truth
- No pre-learning and testing, it just happens
- Black Box

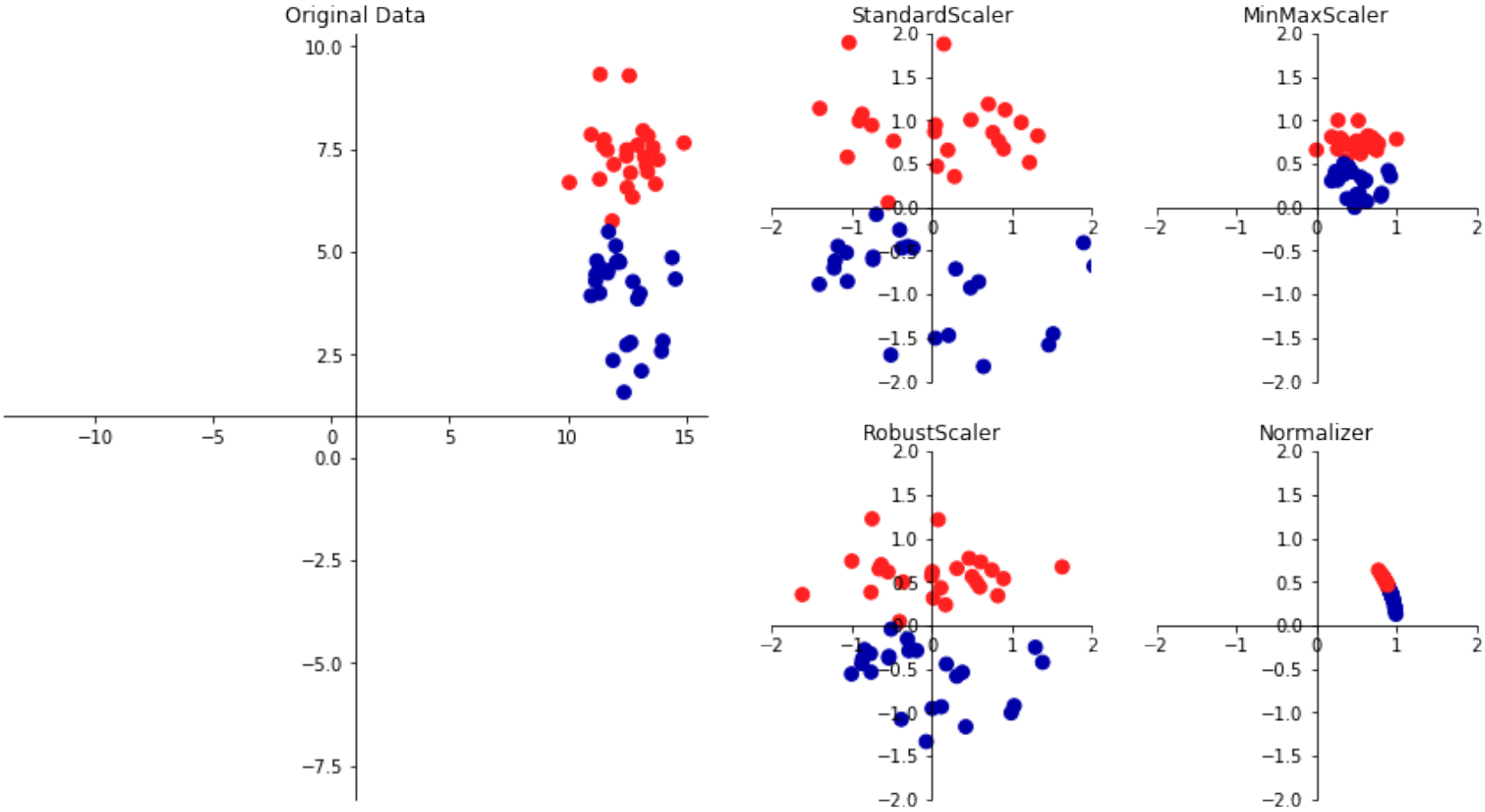
What for?

1. Dimensionality reduction (many features to fewer features)
   - Preprocessing
   - Feature engineering
2. Clustering

Scaling

- Necessary for some supervised machine learning methods (SVM, Neural Networks)
- Four scaling methods
  1. `StandardScaler`: Standarization (mean 0 and variance 1)
  2. `MinMaxScaler`: Features shifted to be between 0 and 1
  3. `RobustScaler`: Normalisation using mean and quartile
  4. `Normalizer`: Projection on unit circle

Scaling, cont.

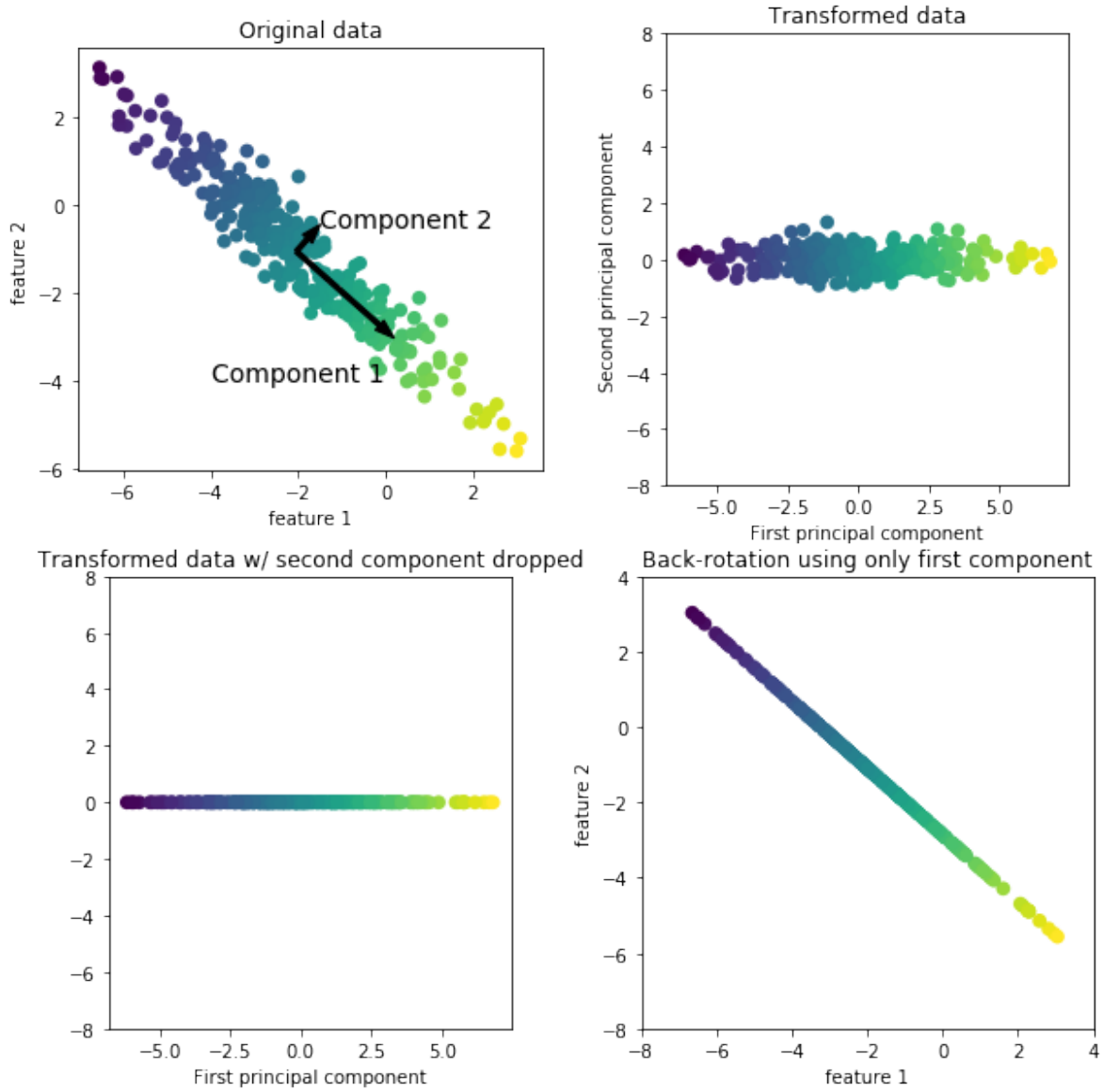


from: Andreas Müller and Sarah Guido (2016): Introduction to Machine Learning with Python, O'Reilly

Principal Component Analysis

- Projection of the entire dataset onto a different feature sub-space
- Extract Principal Components that
  a) represent combinations of all features
  b) are orthogonal to each other
- Components are chosen such that along each axes variance is maximized
- Components are just Eigenvectors belonging to the Eigenvalues of the Covariance Matrix (or the Correlation Matrix)
- Data is usually scaled

Principal Component Analysis, cont.



from: Andreas Müller and Sarah Guido (2016): Introduction to Machine Learning with Python, O'Reilly

Non-Negative Matrix Factorization

- Where PCA required orthogonal components, NMF requires components and coefficients to be non-negative
- Features *need* to be non-negative
- All components equal
- Random initialization leads to non-reproducible outcomes
- Common in Text Mining