W4995 Applied Machine Learning

Dimensionality Reduction

(PCA)

03/25/19

Andreas C. Müller

(Adapted and modified for CC 6021236 @ PCC/Ciencias/UCV by Eugenio Scalise, October 2019)

Unsupervized Learning

Unsupervized Learning

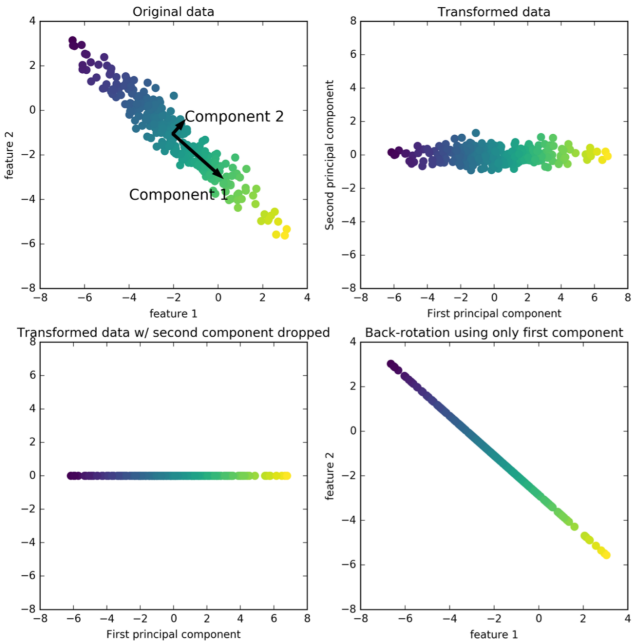
- Unsupervised learning subsumes all kinds of machine learning where there is no known output, no teacher to instruct the learning algorithm.
- Unsupervised learning involves tasks that operate on datasets without labeled responses or target values
- The learning algorithm is just shown the input data and asked to extract knowledge from this data.
- We will look into two kinds of unsupervised learning in this course: **transformations of the dataset** (i.e. dimensionality reduction) and **clustering**.

Applications of Unsupervized Learning

- Visualize structure of a complex dataset.
- Density estimation to predict probabilities of events.
- Compress and summarize the data.
- Extract features for supervised learning.
- Discover important clusters or outliers.

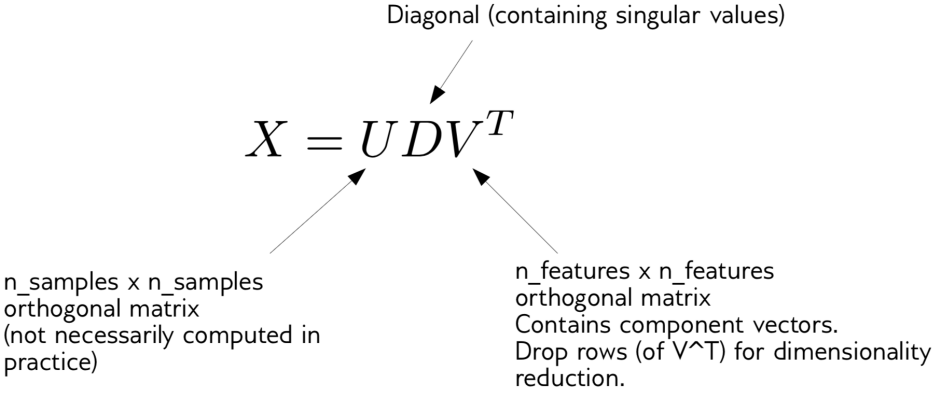
Principal Component Analysis (PCA)

- PCA finds the directions of the maximum variant in the data.
- PCA returns an orthogonal basis of the space where the components are ordered by how much variance of the data they cover and so we get a new basis of the original space.
- This new basis corresponds to rotating the input space (in the example the first component is my X-axis and the second component is my Y-axis and this is the transformation that PCA learns).
- We can use PCA for dimensionality reduction dropping some of the dimensions (they are ordered by decreasing variance).
- The idea is that does a PCA projection covers as much as possible of the data set.
- You can also look at PCA as a de-noising algorithm.

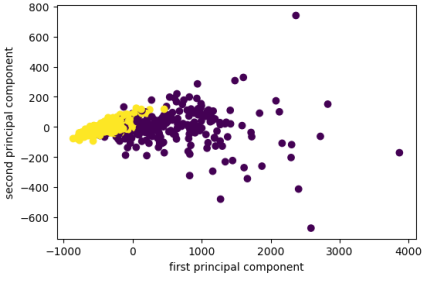

Original data
Component 2
Component 1
feature 2
feature 1
Transformed data
Second principal component
First principal component
Transformed data w/ second component dropped
First principal component
Back-rotation using only first component
feature 2
feature 1

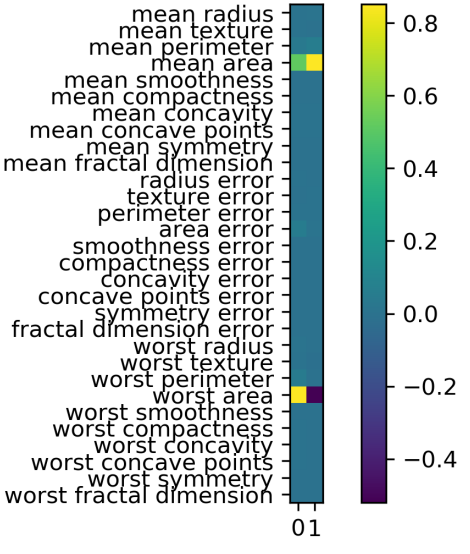
PCA Computation

- Center X (subtract mean).
- In practice: Also scale.
- Compute singular value decomposition (SVD):

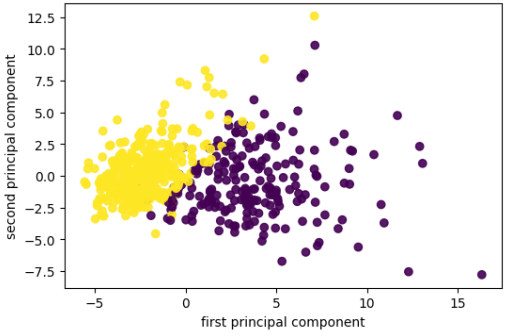


PCA for Visualization

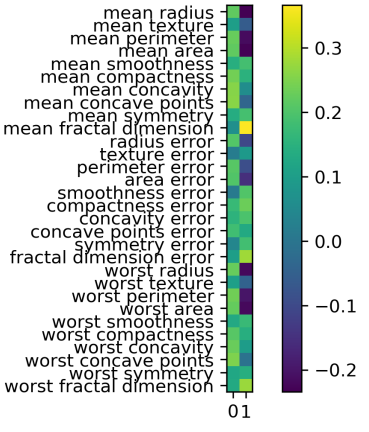


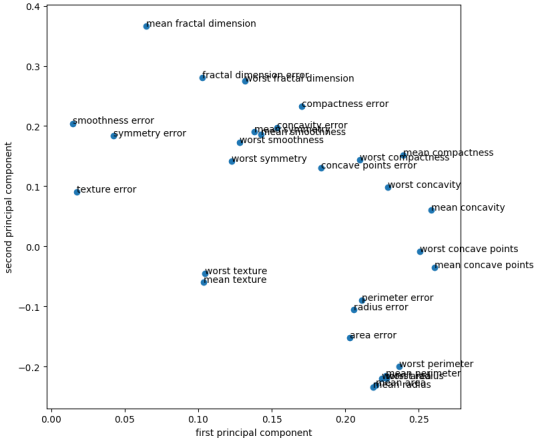


Scaling!


12.5
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
second principal component
−5
0
5
10
15
first principal component

Inspecting components





PCA for regularization

Summary

- PCA good for visualization, exploring correlations
- PCA can sometimes help with classification as regularization or for feature extraction.
- Check LDA and t-SNE: a supervised alternative to PCA.

Next: Clustering