prVis, a Method for Visualizing Dimension Reduction in R

Tiffany Jiang
Norm Matloff
Robert Tucker
Allan Zhao
University of California, Davis

Symposium on Data Science and Statistics 2019, Seattle
Overview

- Principal Component Analysis (Pearson, 1901)
- T-SNE (van der Maatens and Hinton, 2008)
- UMAP (McInnes, Healy, Melville, 2018)
- Diffusion Maps (Coifman, 2006)
- Kernel PCA (Sidhu GS, Asgarian N, Greiner R and Brown MRG, 2012)
- prVis!
Motivation

Swiss Roll Data set

Created to test dimensional reduction.
"Create several points in 2d, map to 3d with a smooth function, and then use an algorithm to map back to 2D"
Swiss Roll, smoothed
Motivation

Principal Component Analysis, Swiss Roll

You don’t really see the four componenta here....
**Motivation**

Uniform Manifold Approximation and Projection for Dimension Reduction, UMAP, Swiss Roll, package uwot

Nor are the components here...
Motivation

tsne, Swiss Roll, package rtsne

Not with t-sne, either...
prVis, a Method for Visualizing Dimension Reduction in R

Tiffany Jiang
Norm Matloff
Robert Tucker
Allan Zhao
University of California, Davis

Motivation

prVis, Swiss Roll

You see the four components with prVis, without color....
prVis, a Method for Visualizing Dimension Reduction in R

Motivation

prVis, Swiss Roll, with color

And even better with color!
prVis

Tiffany Jiang
Norm Matloff
Robert Tucker
Allan Zhao
University of California, Davis

https://github.com/matloff/prVis

• Poly-expansion, and then applies PCA
prVis, a Method for Visualizing Dimension Reduction in R

Tiffany Jiang
Norm Matloff
Robert Tucker
Allan Zhao
University of California, Davis

https://github.com/matloff/prVis

- Poly-expansion, and then applies PCA
- Gnanadesikan and Wilk, 1969
prVis, a Method for Visualizing Dimension Reduction in R

Tiffany Jiang
Norm Matloff
Robert Tucker
Allan Zhao
University of California, Davis

https://github.com/matloff/prVis

- Poly-expansion, and then applies PCA
- Gnanadesikan and Wilk, 1969
- Captures the non-linearity relationship in the data set
prVis, a Method for Visualizing Dimension Reduction in R

Tiffany Jiang
Norm Matloff
Robert Tucker
Allan Zhao
University of California, Davis

https://github.com/matloff/prVis

- Poly-expansion, and then applies PCA
- Gnanadesikan and Wilk, 1969
- Captures the non-linearity relationship in the data set
- Simple!
Classical PCA, Pearson

Positives (Frisvad)

• Works efficiently on large data sets (both in objects and variables)
• Does not assume the multivariate normal, can be applied to all data sets

Negatives

• Not designed to handle non-linear data sets
• Reduces information
t-SNE (van der Maaten and Hinton)

Preserves distance relations

**Positives**

- Works well with non-linear data
- Works well for visualizations

**Negatives**

- Can be inefficient for large data
- Not great for linear data
Here is an example of tsne (left) vs. prVis (right, deg 2) using a dataset from UCI Machine Learning Library. The data set is used to determine whether a star is a pulsar or not.
UMAP (McInnes, Healy, Melville)

Positives

- Can be used for both dimensional reduction and visualization
- Has very fast implementation in multiple programming languages, including Python

Negatives

- Has hyper parameters one has to tune to find a good visualization
Here is an example of UMAP (left) vs. prVis (right, deg 2). Because we know that there are two groups, we wanted to see how clearly separated they were. prVis gives you 2 components and the clear horizontal and vertical groups.
We used diffusion maps on the same swiss roll data set. Not a very good visual here....
We applied the KPCA to the Swiss roll data set, using package kpca with ‘polydot’ as the option. It does similarly, but on other datasets KPCA ran very slowly.
Additional Features

- Support for big memory
- Label row numbers of data points in specific areas of plot
- Built in options for sub sampling
- Outlier removal and outlier removal by category for categorical data
Outlier Removal

prVis, a Method for Visualizing Dimension Reduction in R

Tiffany Jiang
Norm Matloff
Robert Tucker
Allan Zhao
University of California, Davis
prVis, a Method for Visualizing Dimension Reduction in R

Tiffany Jiang
Norm Matloff
Robert Tucker
Allan Zhao
University of California, Davis

Add Row Numbers
prVis, a Method for Visualizing Dimension Reduction in R
Tiffany Jiang
Norm Matloff
Robert Tucker
Allan Zhao
University of California, Davis

https://idyll.pub/post/dimensionality-reduction-293e465c2a3443e8941b016d/
The Beginner’s Guide to Dimensionality Reduction
Matthew Conlen and Fred Hohman
prVis, a Method for Visualizing Dimension Reduction in R

Tiffany Jiang
Norm Matloff
Robert Tucker
Allan Zhao
University of California, Davis

https://github.com/matloff/prVis