Dimensionality Reduction and Interactive Visualization

Zanis Ali Khan (B77407)

Abstract

Getting useful information from high dimensional data is always a challenge is always a matter of discussion but if it is transformed effectively then it becomes very easy and effective way to get useful information from any high dimensionality data. PCA can be used for identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing data.

Introduction

I have taken the Dataset of health insurance market place, which contain multiple files, but I am taking rate file for individual family rates to analysis and plan attributes to identify the plan types. Our rate file contain the rows - 1048576 and dimensions - 24 (which makes rate file = 25, 165, 824). Plan attributes file contains the rows - 77354 and dimensions- 176 (Which makes plan file -13,614,304). Thus, we worked with huge amount of data. The data set is also divided based on year and the data of each year is given.

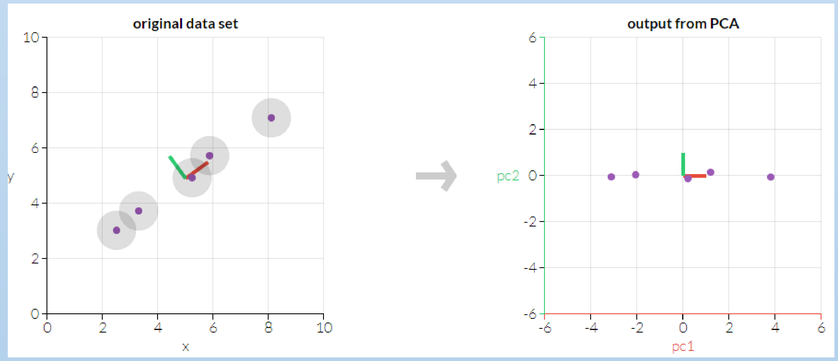
Description

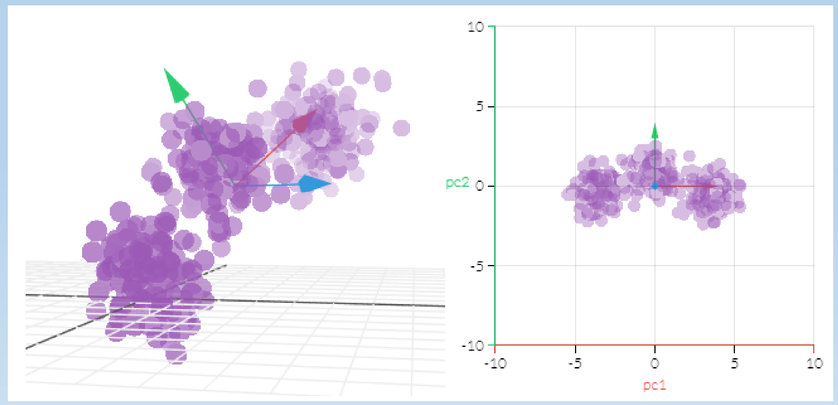
The data, which is used, is from the Centers for Medicare & Medicaid Services (CMS). In the Rate.csv file - This CSV describes the variables contained in the Rate-PUF. Each record relates to one issuer's rates based on plan, geographic rating area, and subscriber eligibility requirements. The Rate PUF is available for plan year 2014, plan year 2015, and plan year 2016

PCA Method Steps.

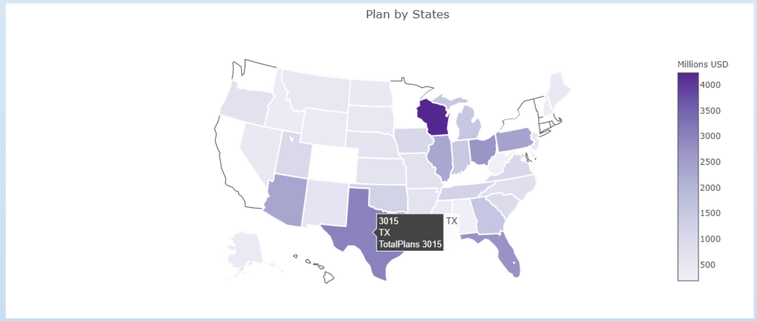
1. Get some data.
2. Subtracting the Mean.
3. Calculating the Co-variance Matrix.
4. Calculate the eigenvectors and eigenvalues of the covariance matrix .
5. Choosing components and forming a feature vector.
6. Deriving the new data set.

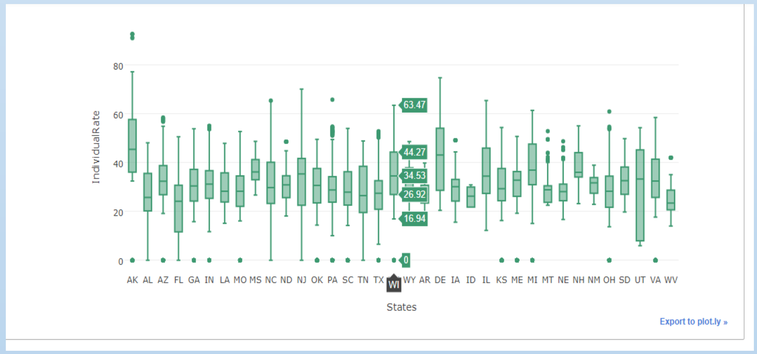
2D and 3D Examples.



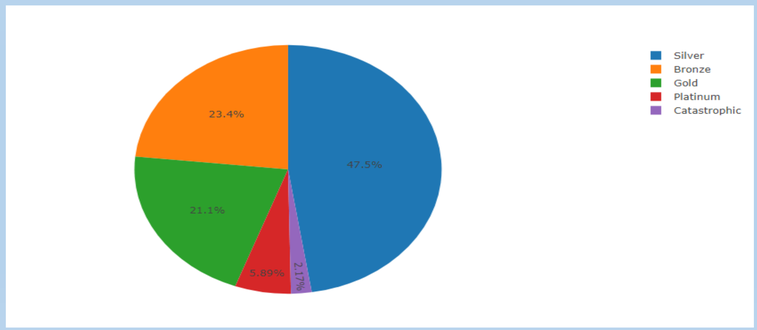


Interactive Graphs





Non-Interactive Graph



References

1. **Dataset** https://www.kaggle.com
2. **PCA Visualization**: http://setosa.io/ev/principal-component-analysis/
3. **PCA** http://www.iro.umontreal.ca/~pift6080/H09/documents/papers/pca_tutorial.pdf

Repository: https://github.com/ZanisAli/AlgorithmicsProjectPoster