

ES 331 Probability and Random Processes 2017-18

Programming Assignment 1 Report

Methodology:

In this Programming Assignment, I have implemented PCA on iris and digits dataset which can be easily imported from sklearn library of python. I can also import it from the source websites but it would be much of a work hence I directly imported it from the sklearn library.

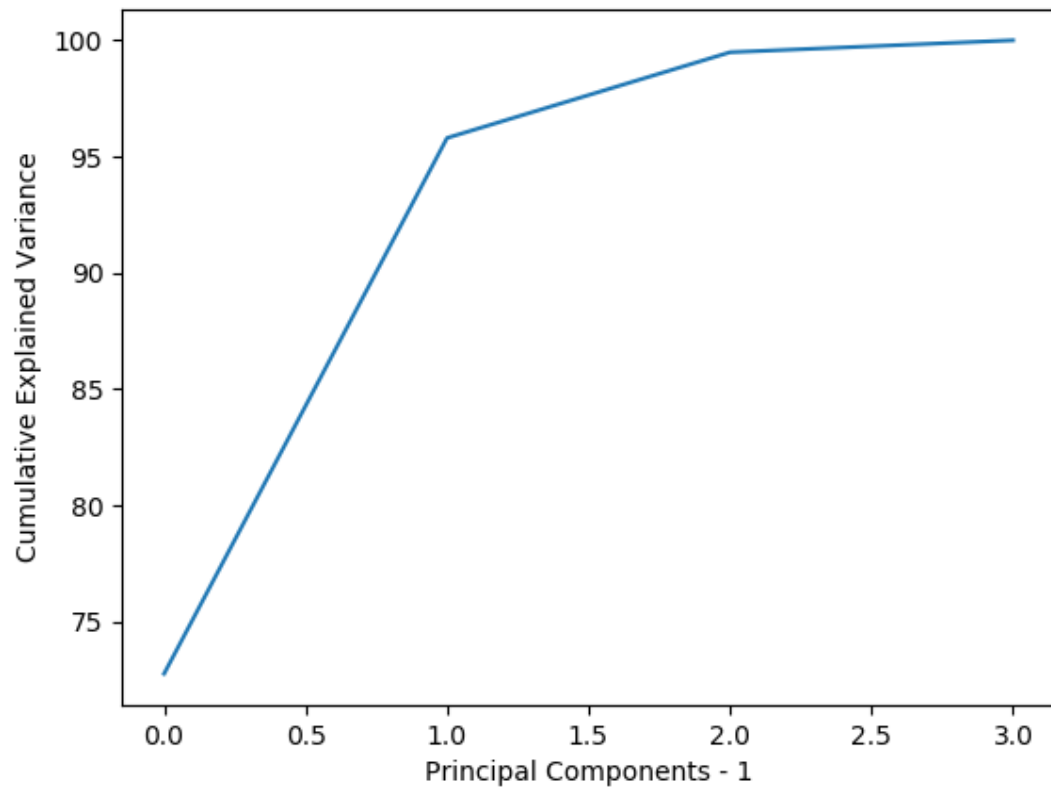
Thereafter from the Data Matrix, I computed the Mean Data Matrix after standardizing the values in the Data Matrix. After which I computed the Covariance Matrix and hence the Eigen Vectors and Eigen Values of the Covariance Matrix.

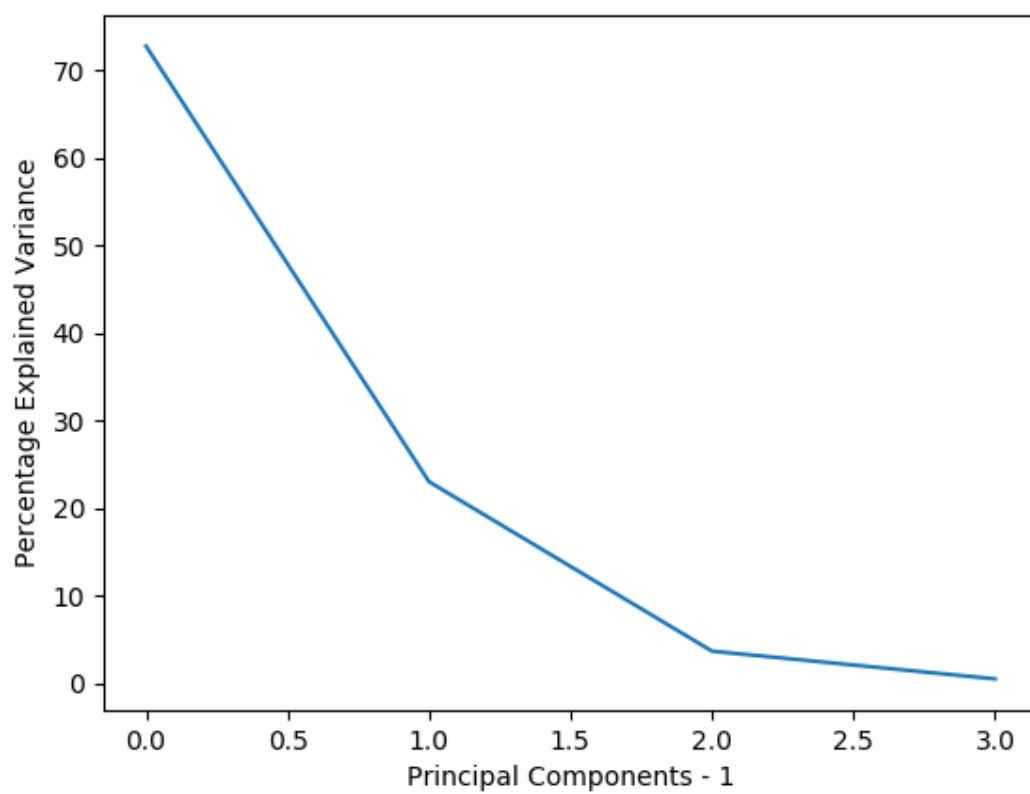
Arranging this Eigen Values in descending order and removing the principal components with the least variances we end up with a reduced Eigen Matrix comprising of Eigen Vectors.

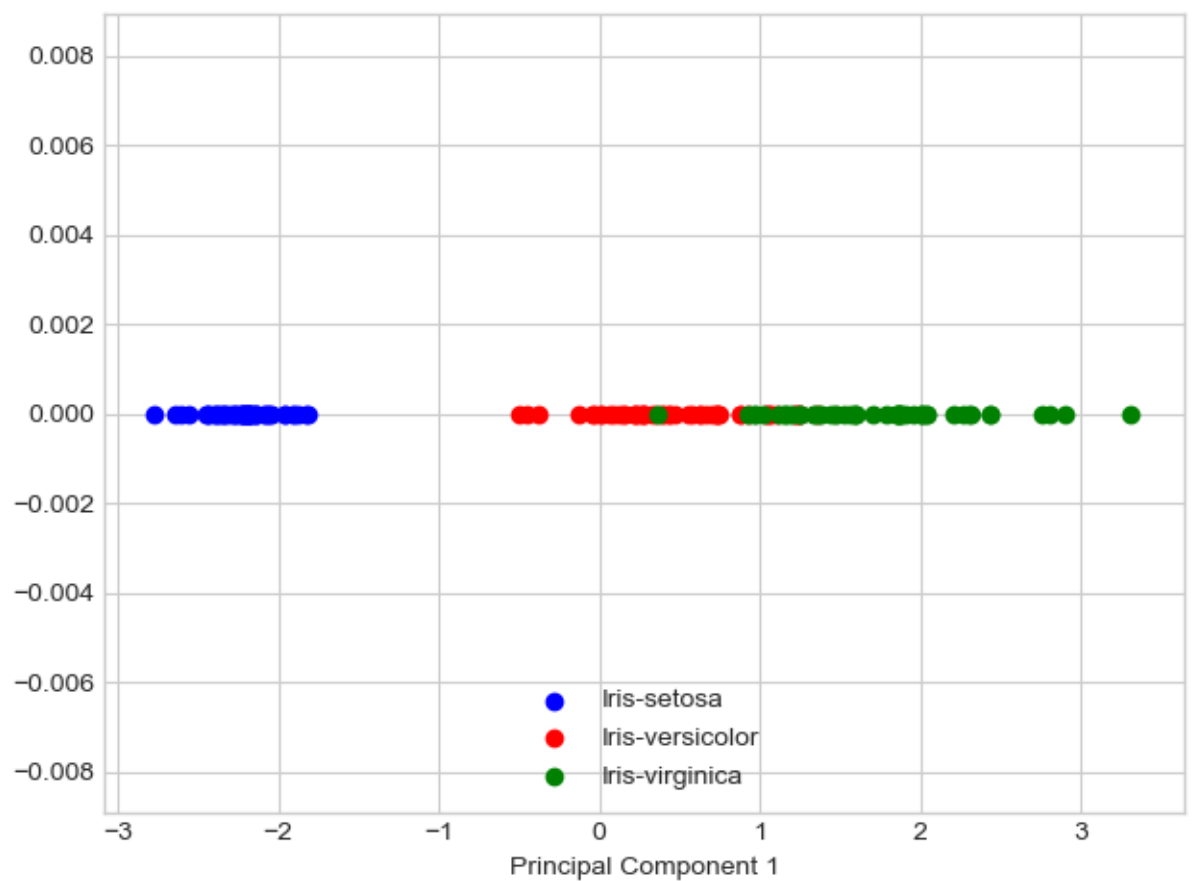
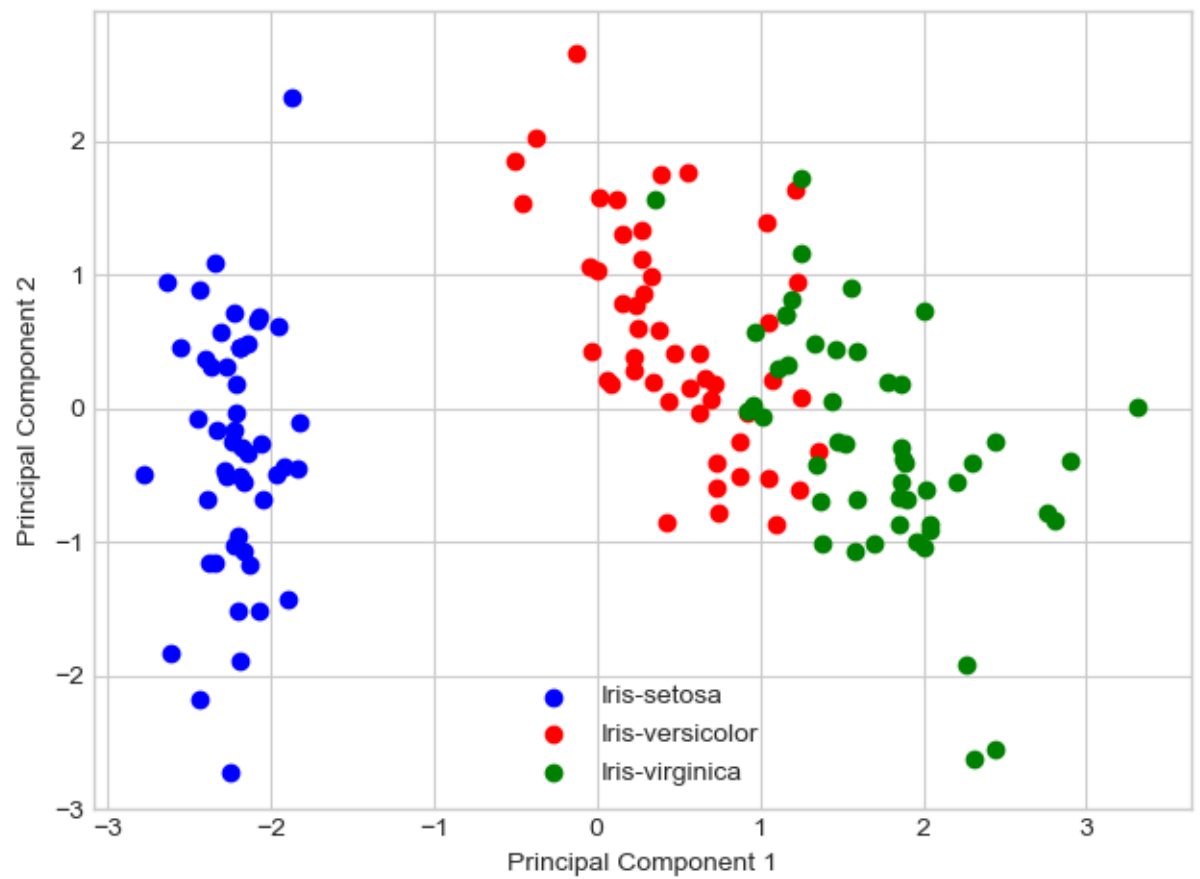
This Eigen Matrix when multiplied with the original data matrix results into Reduced Dimensions Data Matrix which can then be used for different applications.

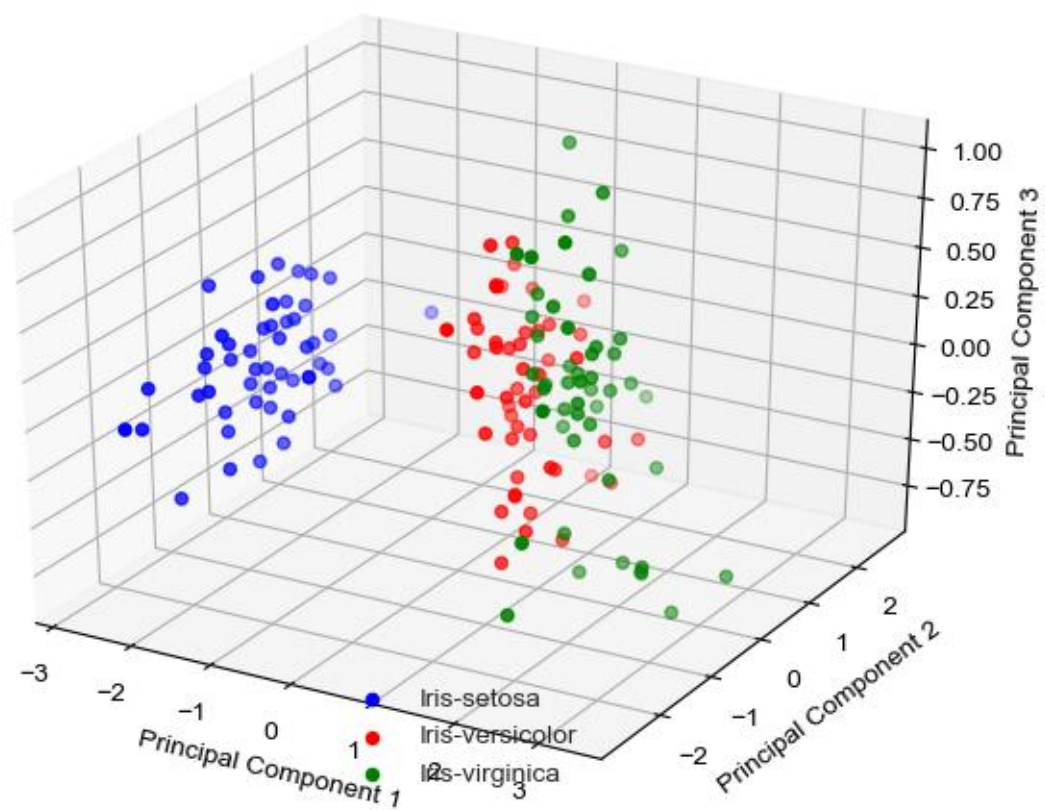
Results:

Iris Dataset

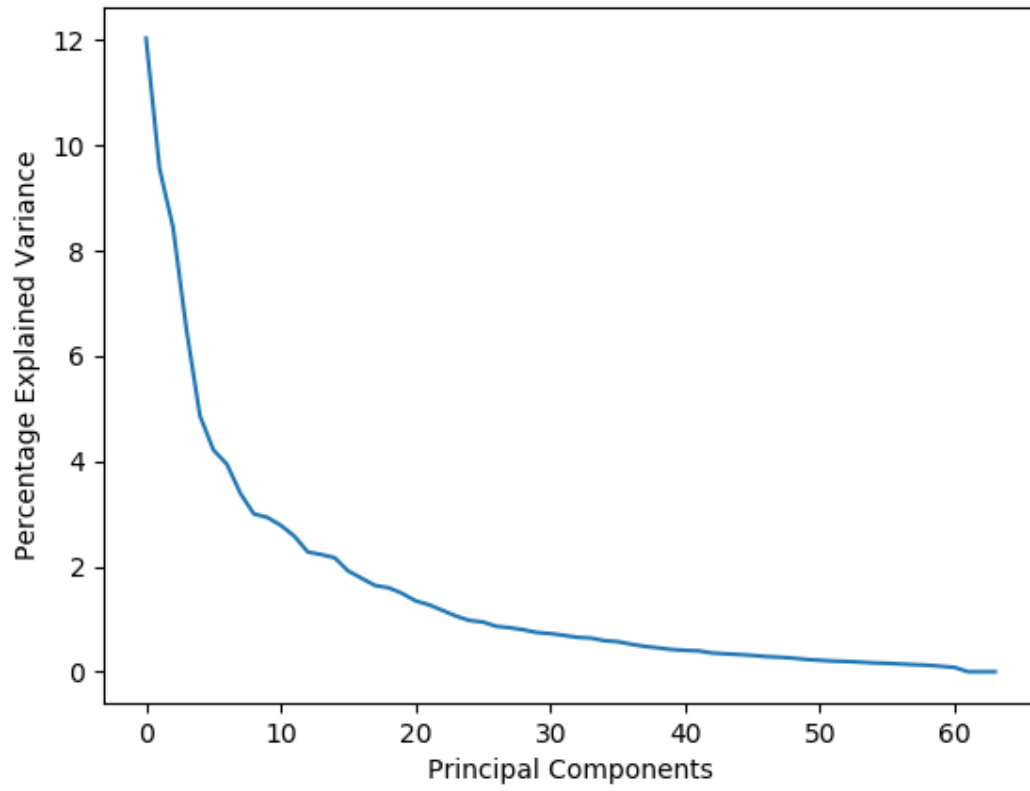


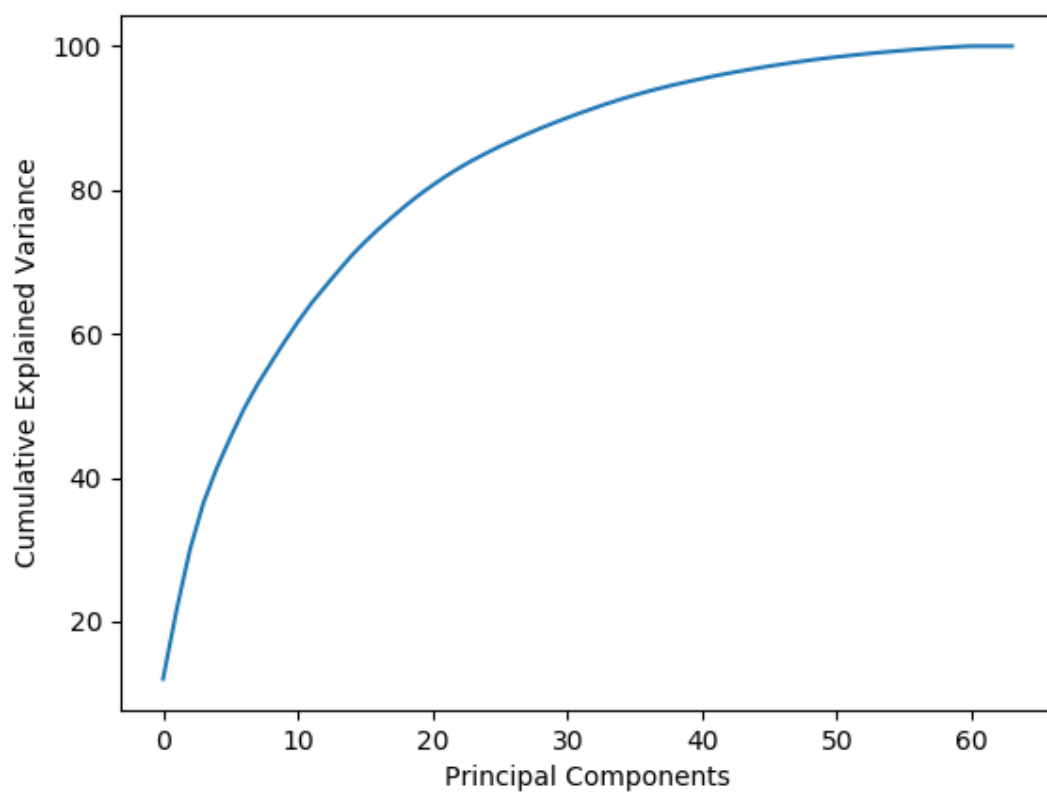


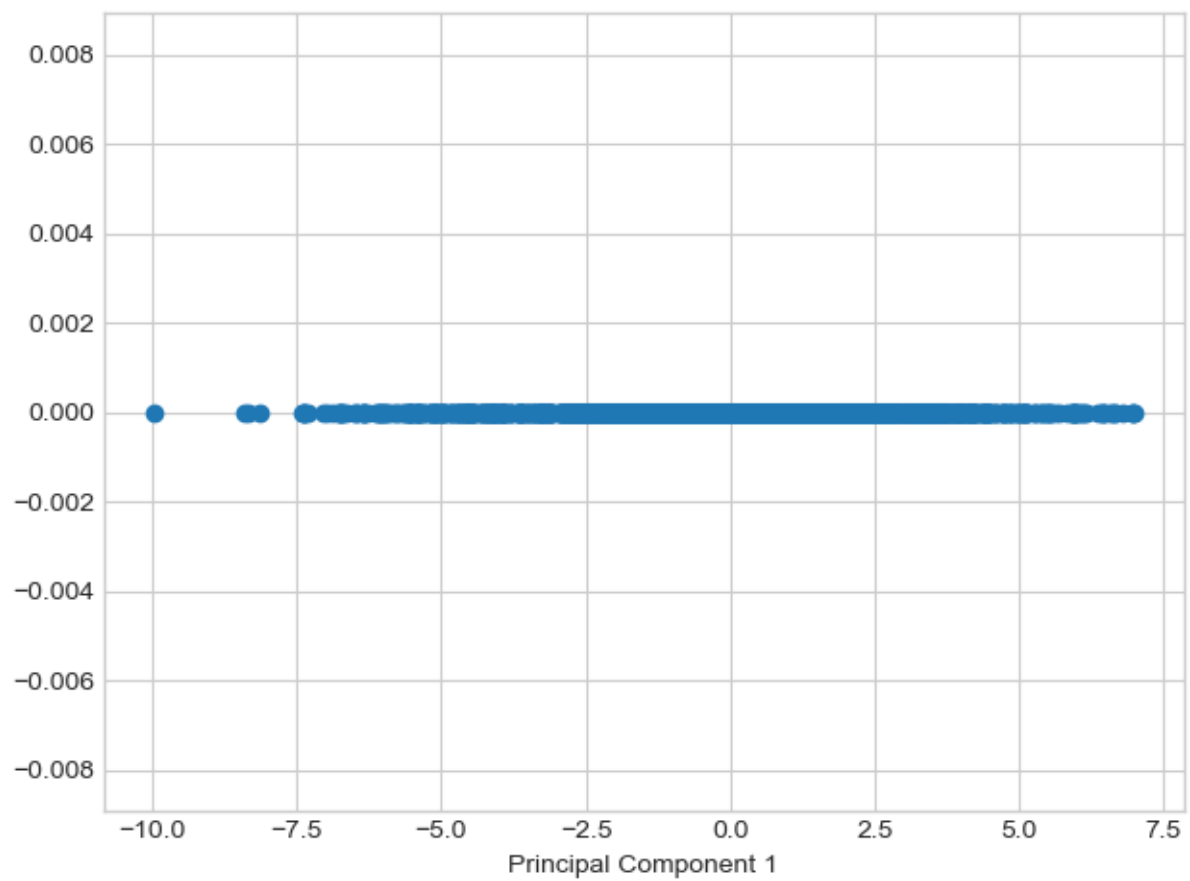


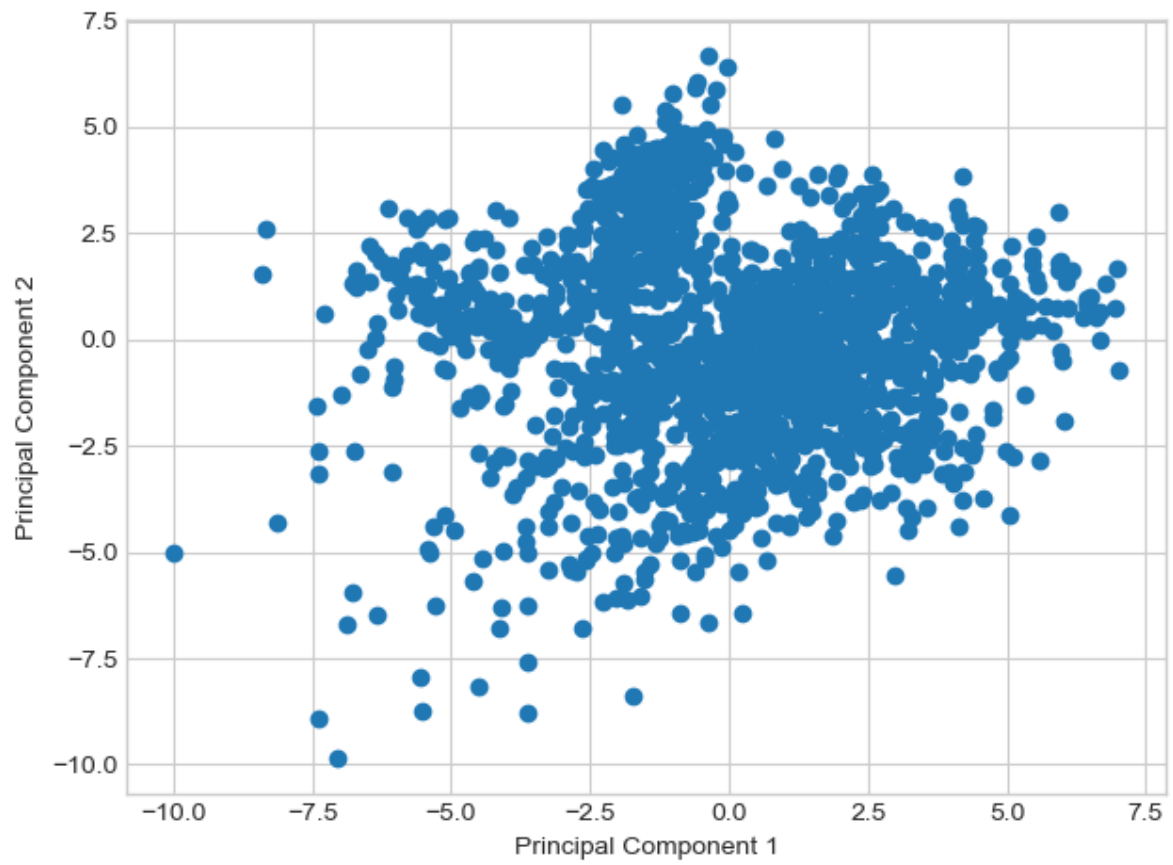


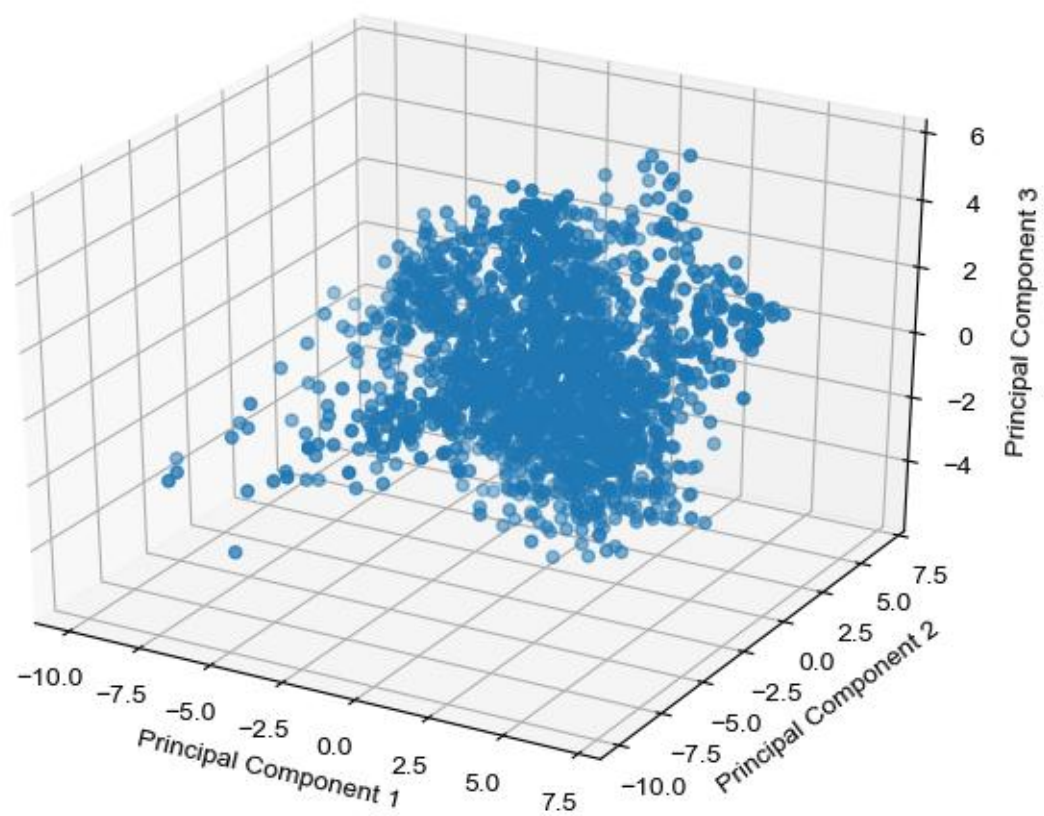
Digits Dataset











Dependencies:

- numpy
- matplotlib
- sklearn for importing datasets (and not performing PCA)
- Rest of the dependencies which are added in the python files are not necessary.

References:

- “Plotting 4d-Data.” *Python - Plotting 4d-Data - Stack Overflow*, stackoverflow.com/questions/31806254/plotting-4d-data.
- “How to Declare Array of Zeros in Python (or an Array of a Certain Size).” *How to Declare Array of Zeros in Python (or an Array of a Certain Size) - Stack Overflow*, stackoverflow.com/questions/4056768/how-to-declare-array-of-zeros-in-python-or-an-array-of-a-certain-size.
- “Principal Component Analysis.” *Wikipedia*, Wikimedia Foundation, 12 Oct. 2017, en.wikipedia.org/wiki/Principal_component_analysis#Covariances.
- “Principal Component Analysis.” *Sebastian Raschka's Website*, 27 Jan. 2015, sebastianraschka.com/Articles/2015_pca_in_3_steps.html.
- “Singular Value.” *Wikipedia*, Wikimedia Foundation, 18 Jan. 2017, en.wikipedia.org/wiki/Singular_value.
- “Numpy.linalg.eig¶.” *Numpy.linalg.eig — NumPy v1.13 Manual*, docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.linalg.eig.html.
- “Scikit-Learn PCA .fit_transform Shape Is Inconsistent (n_samples << m_attributes).” *Machine Learning - Scikit-Learn PCA .fit_transform Shape Is Inconsistent (n_samples << m_attributes) - Stack Overflow*, stackoverflow.com/questions/38276047/scikit-learn-pca-fit-transform-shape-is-inconsistent-n-samples-m-attributes.
- “The Iris Dataset¶.” *The Iris Dataset — Scikit-Learn 0.19.0 Documentation*, scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html.
- Shlens, Jonathon. “A Tutorial on Principal Component Analysis.” [1404.1100] *A Tutorial on Principal Component Analysis*, 3 Apr. 2014, arxiv.org/abs/1404.1100

