

Dimensionality Reduction Techniques and Their Use in Data Visualization

Ilya Ilyankou

In my project, I explored the possibility of visualizing high dimensional data in 3D. I studied and compared popular dimensionality reduction algorithms—*Principal Component Analysis* (PCA), *Multidimensional Scaling* (MDS), and *Sammon Mapping*—and their ability to “shrink” any dataset into three dimensions while preserving as much information about the data as possible. As a result, I created a web application that can take any .csv dataset, analyze it, fill in missing values, and apply one of the above mentioned algorithms. By reducing the number of parameters from many ($N > 3$) to 3, values of three such parameters can be treated as x, y, and z coordinates, allowing each data point to be plotted in 3D. The application makes use of WebGL (Three.js JavaScript library) to create an interactive visualization where one can zoom and rotate the camera view. Through such interaction, users can better understand how data points are related in space, and find hidden patterns in the dataset. The computations are performed in the back-end using Python’s numpy and other libraries.