Laboratory Work 9: PCA and SVD Analysis Report

Rossokha D.V.[1,*]

[1]Department of Computer Science, National Technical University "Kharkiv Polytechnic Institute", Ukraine
*contact: dariia.rossokha@infiz.khpi.edu.ua

This report summarizes the results of Laboratory Work 9, focusing on the application of Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) for dimensionality reduction and data visualization. The analysis was performed using Python's `sklearn.decomposition.PCA` class and related tools to explore the dataset's structure.

The dataset was visualized in 2D and 3D spaces using PCA, with points colored by the 'Channel' feature. The resulting scatter plots, shown in Figure 1, reveal the data's clustering patterns across the first two and three principal components, respectively. SVD was applied to decompose the dataset, and the eigenvalues, calculated as $\lambda_i = \sigma_i^2/(m-1)$, were plotted in descending order (Figure 2). The rapid decline in eigenvalues indicates that the initial components capture most of the data's variance.

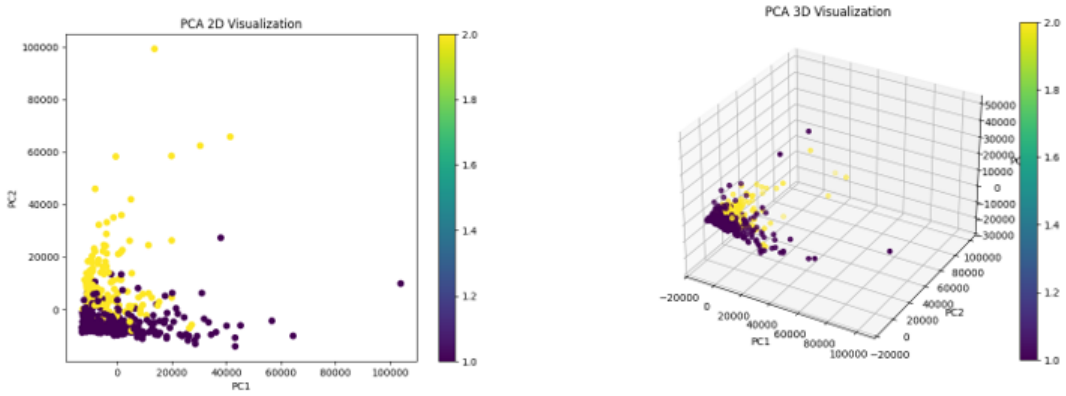

Figure 1: PCA 2D and 3D Visualization.

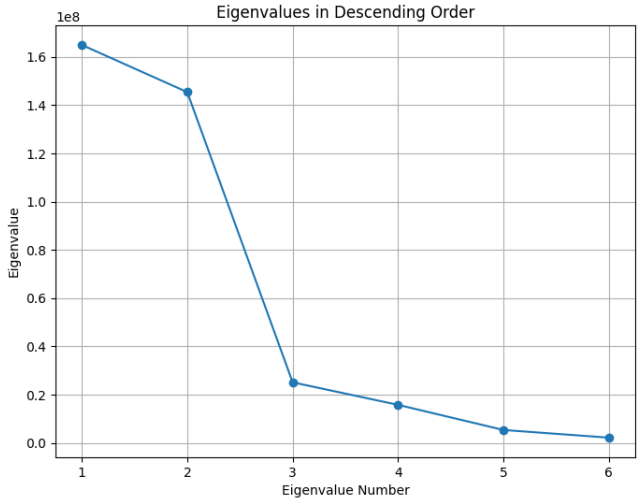

Figure 2: Eigenvalues in Descending Order.

The smallest dimension retaining 80% of the variance was determined to be $d = 2$, with an explained variance of 0.8648, surpassing the threshold (Figure 5). Dimensionality reduction was performed by setting singular values to zero for $i \geq d$ (from the third component onward for $d = 2$). The reverse transformation yielded a Mean Squared Error (MSE) of 8071660.7439 compared to the original data, with a smoothed trend observed in the comparison of the first feature (Figure 3).

Reconstructed data for $d = 2$ and $d = 3$ were visualized using the first d columns, as shown in Figure 4.

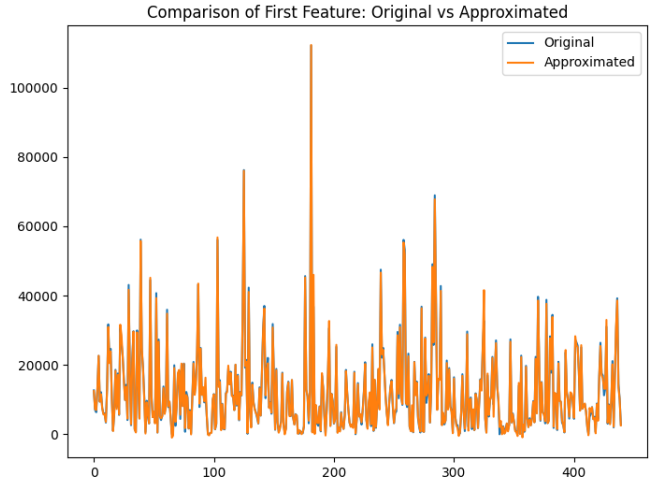

Figure 3: Comparison of First Feature: Original vs Approximated.

These plots retain the general clustering observed in the PCA visualizations but reflect the effects of reduced dimensionality.

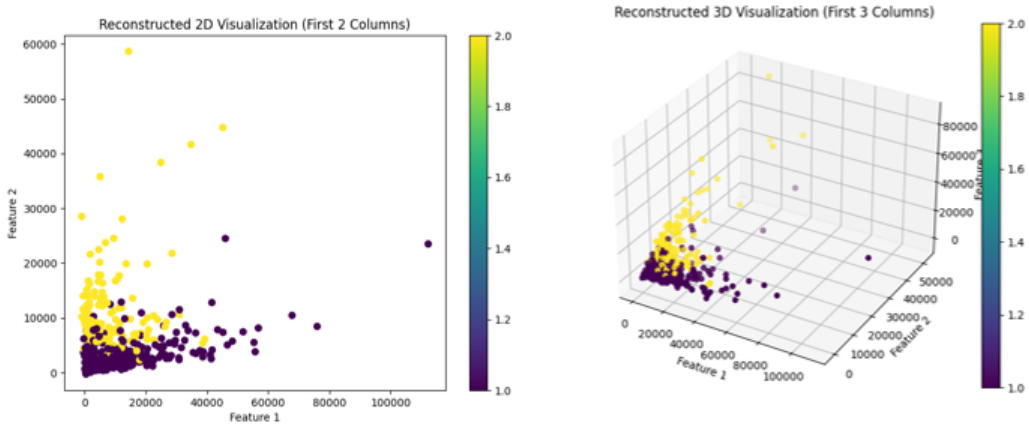

Figure 4: Reconstructed 2D (First 2 Columns) and 3D Visualization (First 3 Columns).

Comment: The analysis successfully demonstrates the effectiveness of PCA and SVD for dimensionality reduction, with two dimensions capturing over 86% of the variance, as evidenced by the explained variance of 0.8648. The visualizations and MSE results confirm that the reduced representations preserve the dataset's core structure while simplifying its complexity.

```
PCA visualization in 2D and 3D
Plot eigenvalues in descending order
Find smallest d for 80% significance
Smallest d: 2 (explained variance: 0.8648)
Reconstruct with d=2 and compare
Mean Squared Error between original and approximated: 8071660.7439
 For d=2
Compare to PCA 2D: Differences may occur due to reconstruction approximation.
For d=3
Compare to PCA 3D: Differences may occur due to reconstruction approximation.
```

Figure 5: Summary of PCA and SVD Dimensionality Reduction Analysis.