ALADS - Assignment - 5

Chenglong Li

May 2024

1 1

Include a plot of the singular values in your code. See Figure 1.

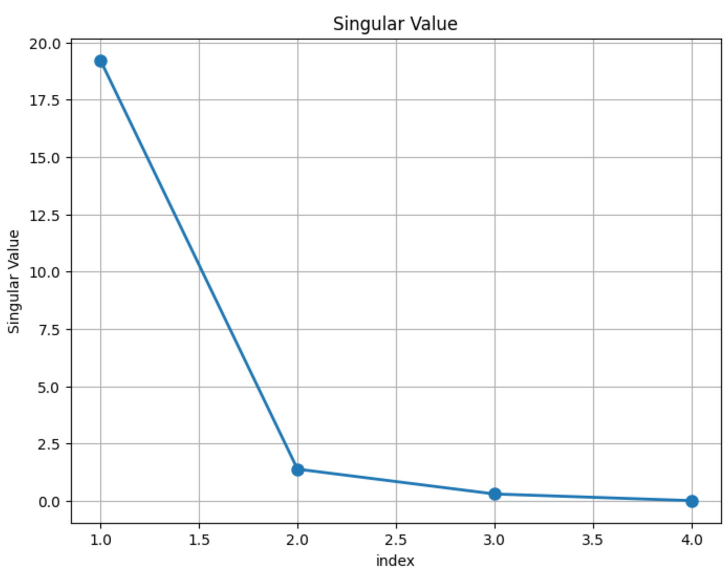

Figure 1: Sigular Values

1.1 1.a

Percent of the total variance is covered by the first eigenvalue= 91.96%. See Figure 2

1.2 1.b

The principal component associated with the largest eigenvalue indicates the direction of the largest variance in the data. This means that it is the most

```
Eigenvalues:
[1.92202892e+01 1.37960341e+00 2.95217138e-01 4.89024219e-03]

Variance Fractions:
[9.19631063e-01 6.60097325e-02 1.41252219e-02 2.33982880e-04]

Percentage of total variance covered by the first eigenvalue: 91.96311%
```

Figure 2: Output: 1.a

important component, capturing the most information (variance) in the data set.

Variables with similar coefficients in the largest eigenvector are highly correlated. Variables with different magnitudes or opposite signs in the eigenvectors are less correlated or negatively correlated. See Figure 3

```
PC1:
[[ 3.94645844]
 [-4.53698958]
 [ 1.76612071]
 [ 3.70825924]
 [-4.88384882]]

 Eigenvalues:
[1.92202892e+01 1.37960341e+00 2.95217138e-01 4.89024219e-03]

 Eigenvectos:
[[ 0.50773119  0.32963362  0.56347395 -0.56218131]
 [ 0.38829179  0.73504644 -0.43622104  0.34445236]
 [-0.54916373  0.46697161  0.58788896  0.36707394]
 [ 0.53838436 -0.36467235  0.38287711  0.65617183]]
```

Figure 3: Output: 1.b

2 2

We start by filling in the missing values in the data using the mode.

$$\text{Eigenvalues} = \begin{bmatrix} 2001.49 & 555.71 & 152.80 & 125.52 & 46.04 & 13.89 & 0.21 \end{bmatrix}$$

Their fraction of the total variance:

$$\text{Fraction of the total variance} = \begin{bmatrix} 69.12\% & 19.19\% & 5.28\% & 4.33\% & 1.59\% & 0.48\% & 0.01\% \end{bmatrix}$$

Plot the principal components in 3D. See Figure 4. Plot the principal components in 2D (principal component 1 versus 2). See Figure 5. Plot the principal components in 2D (principal component 1 versus 3). See Figure 6.

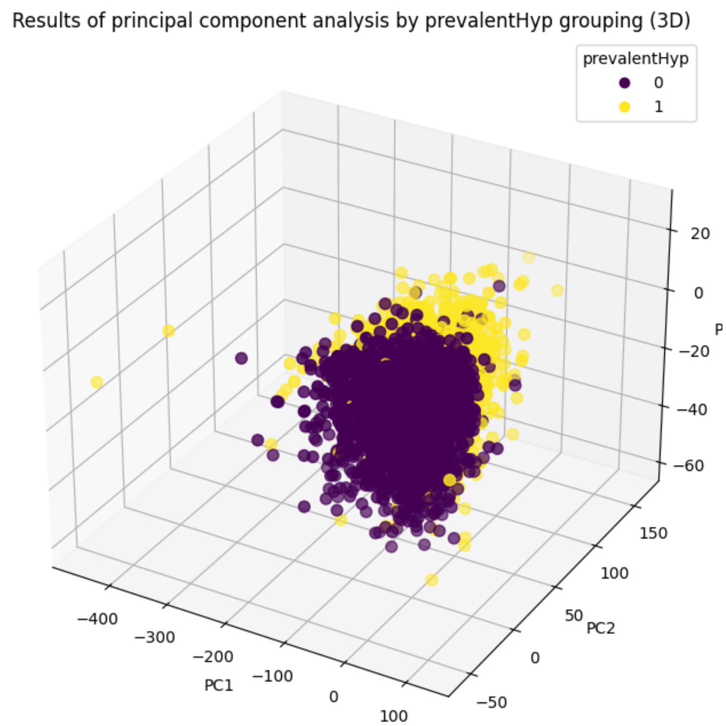

Figure 4: Output: The principal components in 3D

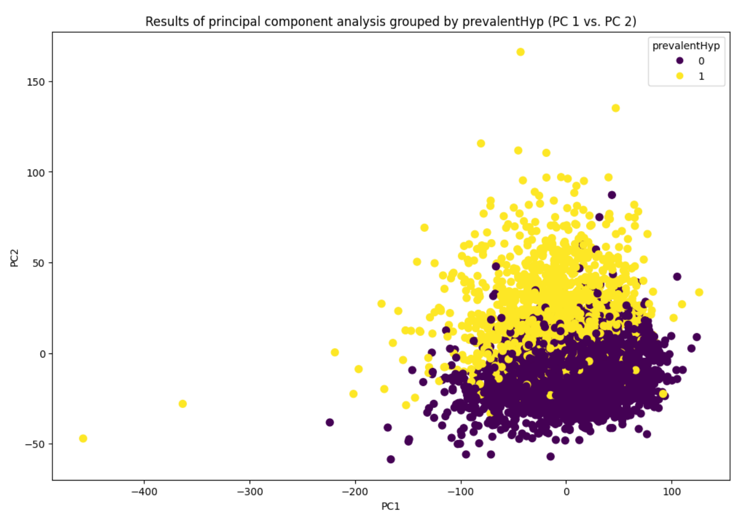

Figure 5: Output: The principal components in 2D(PC1 vs. PC2)

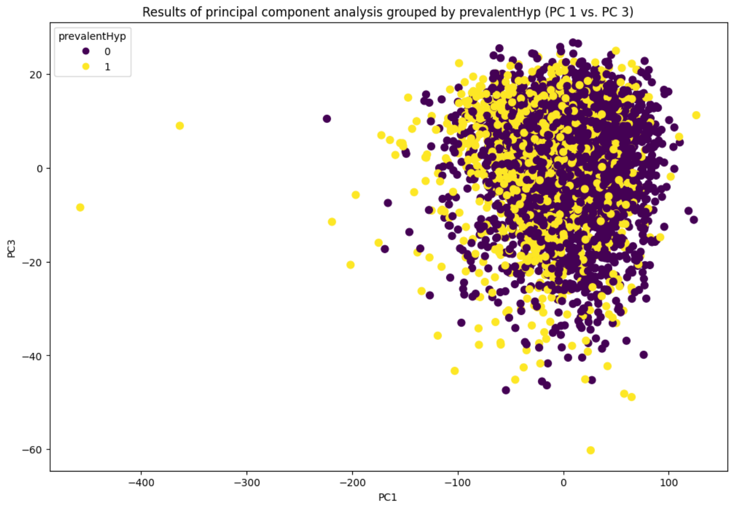

Figure 6: Output: The principal components in 2D(PC1 vs. PC3)

Principal component 1 (PC1) explained the largest percentage of variance (69.12%), indicating that PC1 captured the most significant pattern of variation in the data. This was followed by principal component 2 (19.19%) and principal component 3 (5.28%).

The 3D scatter plot 4 shows the projections of the first three principal components, color-coded by prevalentHyp (prevalent hypertension). This visualization helped us to observe how the data points clustered or separated on the prevalent hypertension grouping.

2D scatter plots show the relationship between principal component 1 and principal component 25, and principal component 1 and principal component 3 6. These charts help us to identify patterns and groupings in the data.

Fisrt eigenvector:

$$\text{eigenvector}_1 = \begin{bmatrix} 0.0007983421 \\ -0.0002172648 \\ -0.0062277357 \\ -0.0136062946 \\ 0.0104078765 \\ -0.0121255441 \\ -0.9997599946 \end{bmatrix}$$

$$Nor(A) \cdot \text{eigenvector}_1 = \text{PC}_1$$

The eigenvector values of the first principal component indicate the contribution of each original variable to principal component 1.

3 Appendix

Code