Temple University

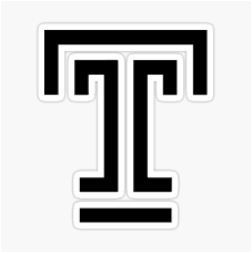
Statistical Science & Data Analytics

The Eigenvectors & Eigenvalues of Principal Component Analysis

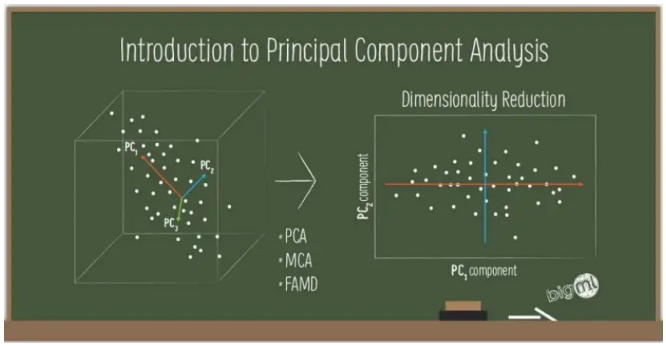
Author
Andrew L. O'Drain

Professor
Dr. Jacqueline Lang

December 04, 2022



Contents



[2]

Figure 1: *Right: variance maximized 3d data projected into 2d*

1 Introduction

Principal component analysis is a statistical technique that is used to reduce the dimensionality of large data sets. It was first mentioned in a paper written by a famous statistician named Karl Pearson in 1901, who once described the technique as quote, "finding lines and planes of closest fit to systems of points in space".[1] In this paper I will only be discussing the mathematical core of PCA, which allows us, with the help of matrix algebra, to find these lines and planes of best fit.

2 Understanding Principal Component Analysis

If one is not familiar with statistics, understanding what principal component analysis is could be a challenge. Therefore, in this section I will explain what PCA accomplishes. But first, we need to understand what *features* are.

Features are another name for random variables in a data set. In fact, from now on I will be using the words 'features', 'variables' and 'columns' interchangeably. Features could be of the discrete data type, like the number of cars that pass on a road at a certain time. Or, they could be of the continuous data type, like someones salary, or the amount of income tax paid over the course of a year.

When an analyst builds models like multiple logistic regression, or logarithmic regression, they use the sum-total of variance within the features to extract meaningful insights from the data. One can think of the variance within features as the hidden information contained within each of the variables, or columns of the data set. Sometimes, when there is a large number of features, the analyst wishes to measure exactly how impactful each of the variables is in explaining the information contained within the dependent, or Y-variable. Ideally, the analyst wishes to keep one-hundred percent of the meaningful variation within the features, but dispose of those features that contribute little or nothing to their model. This is exactly what PCA accomplishes.

The primary output of PCA is a list of the principal components of the data set, one principal component for each column. Principal components can also be referred to as the eigenvectors of the data set. Therefore, from here on, I will be using the words 'principal components' and 'eigenvectors' interchangeably. They are the new variables that are constructed from linear combinations or, in other words, mixtures of all the original variables such that, most of the variance is compressed into the first variable. For example, a 10 dimensional data set, which is a data set with 10 columns, will yield 10 principal components, but PCA tries to put the maximum possible variance into the first variable, the second most variance in the second variable, and so on, until we can formulate a graph that summarizes the percentage of explained variance by each eigenvector. This graph is called a scree plot, which we will discuss more about later. But for now, the scree plot allows the analyst to choose and discard features with low variance, and keep only those that contribute the most information.

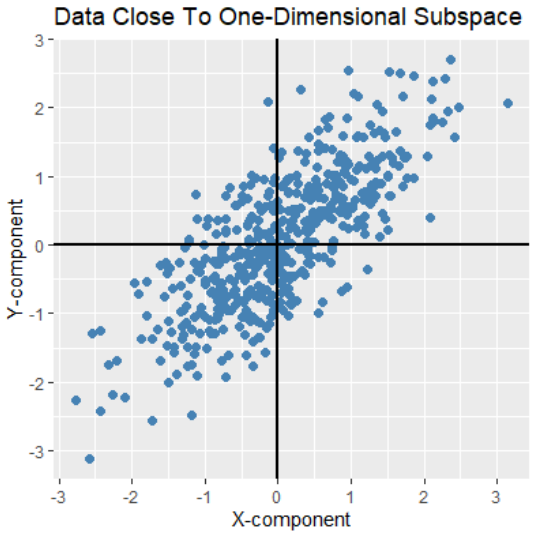


Figure 2: *2-dimensional data that is close to a 1-dimensional subspace we can call C*

In the above diagram we can see an arbitrary 2-dimensional data set that represents the relationship between two arbitrary features with data points that cut through the origin. What if we could preserve all of the information contained within our 2-dimensional subspace and then simplify the problem by projecting the image of the 2-dimensional subspace onto a 1-dimensional subspace, namely, a line? What would that look like? If fact, it would look like a one dimensional line such that the distances between each data point, the variance, would be maximized.

Given the low dimensionality of our example data set, this idea might not seem too useful, however, what if we had a 50-dimensional data set? This is where PCA shines. It allows the analyst to make their data set less crowded with redundant and useless information while preserving most of the useful information. With that being said, the question remains, how can we fit higher dimensional sub-spaces to lower dimensional sub-spaces using linear algebra?

3 Subspace Fitting: Centered Data

Given that PCA appears to be overall complex, the mathematical procedures are straightforward. The task of finding principal components is a bit easier if our data is centered and already passes through the origin.

In the next example, we are going to find the 1-dimensional subspace that best approximates A. We will also find the total squared distance of the data points to the subspace we fit. This may seem much like linear curve fitting, however, it is slightly different. Subspace fitting minimizes the orthogonal distances from each data point to the line (eigenvector), whereas, linear curve fitting minimizes the vertical distances from each data point to the line.

3.1 Finding Eigenvalues & Eigenvectors: Centered 2-Dimensional Data

First we calculate A^TA to get our symmetric matrix, find $det(A-\lambda I)$ and $p_A(\lambda)$, then, derive the roots from $p_A(\lambda)$, the characteristic polynomial. This will yield the eigenvalues. We will then plug in the eigenvalues into $A-\lambda I$ to derive a corresponding eigenspace for every eigenvector that we find. The result will be each eigenvector and eigenvalue of the symmetric matrix. *Consider the following collection of points in* $\mathbb{R}^2$.[6]

$$A = \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \end{bmatrix} \right\}$$

$$A = \begin{bmatrix} 1 & -2 & 3 & 2 & -2 & 3 & -3 & 0 & 0 & 0 \\ -1 & 1 & -2 & -2 & 1 & -1 & 0 & 3 & 0 & -2 \end{bmatrix}$$

$$B = A^TA = \begin{bmatrix} 40 & -18 \\ -18 & 25 \end{bmatrix}$$

$$det(B-\lambda_n I) = \begin{bmatrix} 40-\lambda_n & -18 \\ -18 & 25-\lambda_n \end{bmatrix}$$

$$\begin{aligned} p_B(\lambda) &= (40-\lambda)(25-\lambda) - (-18)(-18) \\ &= \lambda^2 - 65\lambda + 676 \end{aligned}$$

$$\lambda_1 = \frac{-(-65) + \sqrt{(-65^2) - 4(1)(676)}}{2(1)} = 52$$

$$\lambda_2 = \frac{-(-65) - \sqrt{(-65^2) - 4(1)(676)}}{2(1)} = 13$$

$$\bar{v}_1 = \begin{bmatrix} -\frac{3}{2} \\ 1 \end{bmatrix} \qquad \bar{v}_2 = \begin{bmatrix} \frac{2}{3} \\ 1 \end{bmatrix}$$

As we can see, we have two eigenvalues $\{\lambda_1, \lambda_2\}$ and two eigenvectors $\{\bar{v}_1, \bar{v}_2\}$. If we pick the largest eigenvalue, $\lambda_1 = 52$, and its corresponding eigenvector $\bar{v}_1 = (-\frac{3}{2}, 1)$, this is the closest 1-dimensional subspace that is spanned by $\{\bar{v}\}$. In other words, the first eigenvector points to the direction in space in which the data has the largest variance. It essentially tells us how to reorient the initial $\{x, y\}$ axes, such that we get the best view of the data.

3.2 Finding Eigenvalues & Eigenvectors: Centered 3-Dimensional Data

To further demonstrate the finding of lower dimensional sub-spaces, we once again are going to follow the previous procedure. But instead of just finding the closest 1-dimensional subspace, we will also find the closest 2-dimensional subspace, as well as, the closest 3-dimensional subspace. One subspace for each dimension. *Consider the following collection of points in* $\mathbb{R}^3$.[6]

$$A = \left\{ \begin{bmatrix} 0 \\ 0 \\ 9 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ -4 \end{bmatrix}, \begin{bmatrix} 6 \\ -2 \\ 13 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} \right\}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 6 & 1 & 0 & 0 & 2 & 4 & 1 & 0 \\ 0 & 0 & 0 & 5 & -2 & 0 & 2 & 1 & 0 & 0 & -3 & 0 \\ 9 & -3 & 1 & -4 & 13 & 0 & 0 & 1 & 0 & 0 & 0 & 4 \end{bmatrix}$$

$$B = A^T A = \begin{bmatrix} 59 & -15 & 75 \\ -15 & 43 & -45 \\ 75 & -45 & 293 \end{bmatrix}$$

$$det(B - \lambda_n I) = \begin{bmatrix} 59 - \lambda_n & -15 & 75 \\ -15 & 43 - \lambda_n & -45 \\ 75 & -45 & 293 - \lambda_n \end{bmatrix}$$

$$p_B(\lambda) = -\lambda^3 - 395\lambda^2 + 24548\lambda - 417316$$

$$= -(\lambda - 323)(\lambda - 38)(\lambda - 34)$$

$$\lambda_1 = 323 \;\; \lambda_2 = 38 \;\; \lambda_3 = 34$$

$$\bar{v}_1 = \begin{bmatrix} -\frac{5}{17} \\ \frac{3}{17} \\ 1 \end{bmatrix} \qquad \bar{v}_2 = \begin{bmatrix} -\frac{5}{2} \\ \frac{3}{2} \\ 1 \end{bmatrix} \qquad \bar{v}_2 = \begin{bmatrix} \frac{3}{5} \\ 1 \\ 0 \end{bmatrix}$$

3.3 Sum of Squared Distances to a Subspace: Centered 3-Dimensional Data

Now that we have our eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\}$ in decreasing order, along with the associated eigenvectors $\{\bar{v}_1, \bar{v}_2, \bar{v}_3\}$, we can find the closest subspaces in $\{\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3\}$. The closest 1-dimensional subspace is spanned by the eigenvector associated with the largest eigenvalue $\{\bar{v}_1\}$. The closest 2-dimensional subspace is spanned by the eigenvectors associated with the first and second largest eigenvalues $\{\bar{v}_1, \bar{v}_2\}$, and the closest 3-dimensional subspace is spanned by all three eigenvectors $\{\bar{v}_1, \bar{v}_2, \bar{v}_3\}$. Furthermore, we have the total squared distances to each respective subspace.

$$SSd_{\mathbb{R}} = \lambda_2 + \lambda_3 = 323 + 38 = 361$$

$$SSd_{\mathbb{R}^2} = \lambda_3 = 34$$

$$SSd_{\mathbb{R}^3} = 0 \quad \text{since all data points lie in } \mathbb{R}^3$$

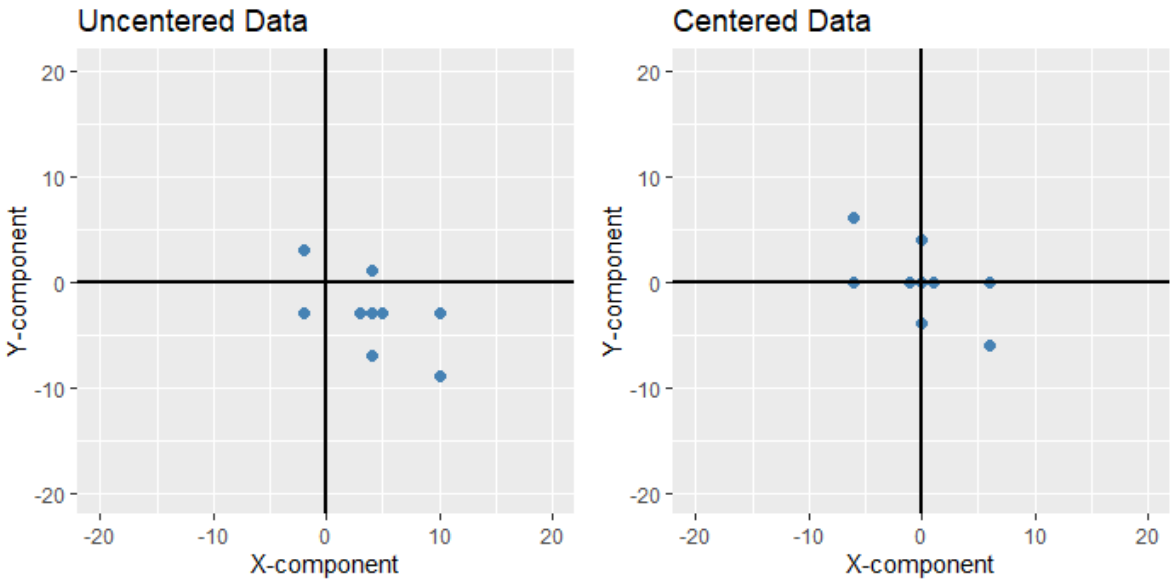


Figure 3: *Left: un-centered data in a 2-dimensional subspace that is close to a 1-dimensional subspace. Right: data after we subtract the mean from every vector in the data set.*

4 Affine Subspace Fitting: Uncentered Data

The above examples are an ideal scenario regarding fitting high dimensional data to lower dimensional subspaces. Most of the time, the data that we gather will not be as cooperative. This is where the affine sub-space fitting problem comes in. The procedure is about the same as in the examples above, however, there is an additional step. We must compute the centroid of the data first. The centroid is the vector in which $\bar{y}$ and $\bar{x}$ are equal to one another. Once we calculate the centroid, we can subtract it from our set of vectors. The resulting set of data points will then be centered around the origin (0, 0). Note, this does nothing to change the position of the data points relative to one another. It only shifts the data so that it is mathematically centered about the origin.

In the next example, I will compute the centroid, as well as, find the 1-dimensional affine subspace that best approximates this collection of points. The last step will be to find the total squared distance of the points to the subspace. *Consider the following collection of points in* $\mathbb{R}^2$. [6]

$$A = \left\{ \begin{bmatrix} 4 \\ -3 \end{bmatrix}, \begin{bmatrix} 10 \\ -9 \end{bmatrix}, \begin{bmatrix} 4 \\ -7 \end{bmatrix}, \begin{bmatrix} -2 \\ 3 \end{bmatrix}, \begin{bmatrix} 10 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ -3 \end{bmatrix}, \begin{bmatrix} 5 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ -3 \end{bmatrix}, \begin{bmatrix} 3 \\ -3 \end{bmatrix} \right\}$$

$$A = \begin{bmatrix} 4 & 10 & 4 & -2 & 10 & 4 & 5 & 4 & -2 & 3 \\ -3 & -9 & -7 & 3 & -3 & -3 & -3 & 1 & -3 & -3 \end{bmatrix}$$

4.1 Compute Centroid & Center Data: Uncentered 2-Dimensional Data

$$B = \frac{1}{10}\left\{\begin{bmatrix}4\\-3\end{bmatrix} + \begin{bmatrix}10\\-9\end{bmatrix} + \begin{bmatrix}4\\-7\end{bmatrix} + \begin{bmatrix}-2\\3\end{bmatrix} + \begin{bmatrix}10\\-3\end{bmatrix} + \begin{bmatrix}4\\-3\end{bmatrix} + \begin{bmatrix}5\\-3\end{bmatrix} + \begin{bmatrix}4\\1\end{bmatrix} + \begin{bmatrix}-2\\-3\end{bmatrix} + \begin{bmatrix}3\\-3\end{bmatrix}\right\}$$

$$= \frac{1}{10}\begin{bmatrix}40\\-30\end{bmatrix} = \begin{bmatrix}4\\-3\end{bmatrix}$$

$$C = A - B = \left\{\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}-6\\6\end{bmatrix}, \begin{bmatrix}0\\4\end{bmatrix}, \begin{bmatrix}6\\-6\end{bmatrix}, \begin{bmatrix}-6\\0\end{bmatrix}, \begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}-1\\0\end{bmatrix}, \begin{bmatrix}0\\-4\end{bmatrix}, \begin{bmatrix}6\\0\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}\right\}$$

4.2 Eigenvalues & Eigenvectors: Uncentered 2-Dimensional Data

$$D = C^T C = \begin{bmatrix}146 & -72\\-72 & 104\end{bmatrix}$$

$$det(D - \lambda_n I) = \begin{bmatrix}146 - \lambda_n & -72\\-72 & 104 - \lambda_n\end{bmatrix}$$

$$p_D(\lambda) = \lambda^2 - 250\lambda + 10000$$

$$\lambda_1 = \frac{-(-250) + \sqrt{(-250^2) - 4(1)(10000)}}{2(1)} = 200$$

$$\lambda_2 = \frac{-(-250) - \sqrt{(-250^2) - 4(1)(10000)}}{2(1)} = 50$$

$$\bar{v}_1 = \begin{bmatrix}-\frac{4}{3}\\1\end{bmatrix} \qquad \bar{v}_2 = \begin{bmatrix}\frac{3}{4}\\1\end{bmatrix}$$

$$B + C = \{b + c \mid c \in C\} = \left\{\begin{bmatrix}4\\-3\end{bmatrix} + y\begin{bmatrix}-\frac{4}{3}\\1\end{bmatrix} \mid y \in \mathbb{R}\right\} \in \mathbb{R}$$

Now that we have both our eigenvectors and their associated values we can find the closest 1-dimensional subspace as well as the sum of the squared distances from it. Remember, when reducing 2-dimensions to 1-dimension the desired subspace W is spanned by the eigenvector with the largest eigenvalue. The total squared distance is the λ_2.

$$SSd_{\mathbb{R}} = \lambda_2 = 50$$

$$SSd_{\mathbb{R}^2} = 0 \quad \text{since all data points lie in } \mathbb{R}^2$$

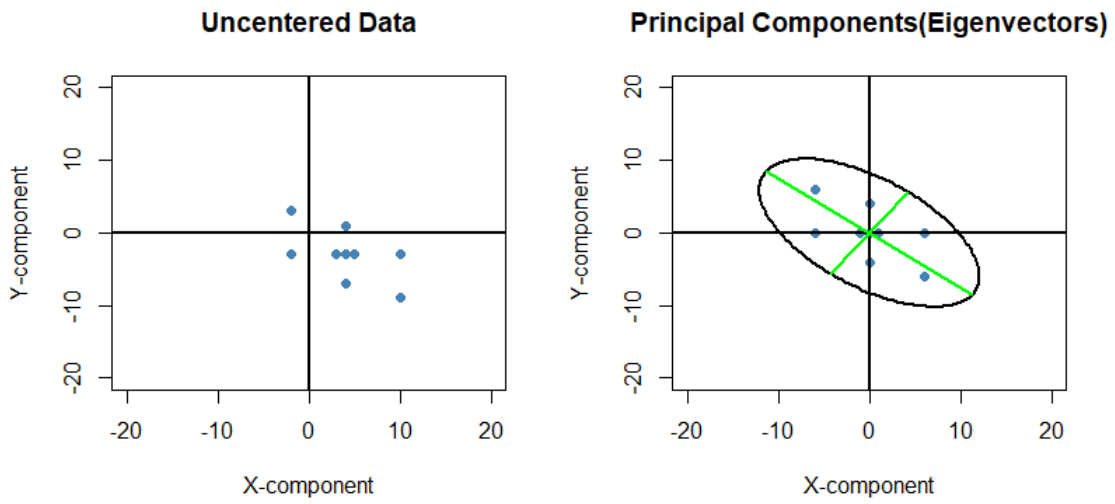


Figure 4: *Left: uncentered data Right: centered data with principal components overlaid.The orthogonal eigenvectors span the direction of greatest variance*

4.3 Tying it All Together

For our final example, we will use three dimensional data to tie all of the concepts together into a complete analysis. We will first visualize our data in three dimensions, compute the centroid, compute our nxn symmetric matrix, then find the closest 1 and 2 dimensional affine subspaces that minimizes the square distances from our data to the subspace. *Consider the following collection of points in* $\mathbb{R}^3$. [6]

$$A = \left\{ \begin{bmatrix} 0 \\ -2 \\ 2 \end{bmatrix}, \begin{bmatrix} 6 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 5 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ -2 \\ -2 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 2 \end{bmatrix} \right\}$$

4.4 Compute Centroid & Center Data: Uncentered 3-Dimensional Data

$$B = \frac{1}{14} \left\{ \begin{bmatrix} 0 \\ -2 \\ 2 \end{bmatrix} + \begin{bmatrix} 6 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \\ 6 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \\ 5 \end{bmatrix} + \begin{bmatrix} 5 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ -2 \\ -2 \end{bmatrix} + \begin{bmatrix} 5 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ -2 \\ 2 \end{bmatrix} \right\}$$

$$= \frac{1}{14} \begin{bmatrix} 42 \\ 0 \\ 28 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}$$

$$C = A - B = \left\{ \begin{bmatrix} -3 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ -4 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 0 \end{bmatrix} \right\}$$

$$C = \begin{bmatrix} -3 & 3 & -1 & 0 & 0 & 2 & 1 & 2 & -2 & 0 & 0 & 0 & -2 & 0 \\ -2 & 2 & 2 & 2 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \\ 0 & 0 & 4 & 0 & 3 & 0 & -4 & 0 & 0 & -3 & -1 & 1 & 0 & 0 \end{bmatrix}$$

4.5 Eigenvalues & Eigenvectors: Uncentered 3-Dimensional Data

$$D = C^T C = \begin{bmatrix} 36 & 8 & -8 \\ 8 & 24 & 16 \\ -8 & 16 & 52 \end{bmatrix}$$

$$det(D - \lambda_n I) = \begin{bmatrix} 36 - \lambda_n & 8 & -8 \\ 8 & 24 - \lambda_n & 16 \\ -8 & 16 & 52 - \lambda_n \end{bmatrix}$$

$$p_D(\lambda) = -\lambda^3 + 112\lambda^2 - 3600\lambda - 28800$$

$$= -(\lambda - 12)(\lambda - 40)(\lambda - 60)$$

$$\lambda_1 = 12 \;\; \lambda_2 = 40 \;\; \lambda_3 = 60$$

$$\bar{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \qquad \bar{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \qquad \bar{v}_3 = \begin{bmatrix} -\frac{1}{5} \\ \frac{2}{5} \\ 1 \end{bmatrix}$$

4.6 Sum of Squared Distance to $\mathbb{R}^n$: Uncentered 3-Dimensional Data

$$SSd_{\mathbb{R}} = \lambda_2 + \lambda_3 = 40 + 60 = 100$$

$$SSd_{\mathbb{R}^2} = \lambda_3 = 60$$

$$SSd_{\mathbb{R}^3} = 0 \quad \text{since all data points lie in } \mathbb{R}^3$$

4.7 Closest One, Two, & Three Dimensional Subspaces

$$B + C = \{b + c \mid c \in C\} = \left\{ \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} + x \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \mid x \in \mathbb{R} \right\} \in \mathbb{R}$$

$$B + C = \{b + c \mid c \in C\} = \left\{ \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} + \left\{ x \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} + y \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \right\} \mid x, y \in \mathbb{R} \right\} \in \mathbb{R}^2$$

$$B + C = \{b + c \mid c \in C\} = \left\{ \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} + \left\{ x \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} + y \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -\frac{1}{5} \\ \frac{2}{5} \\ 1 \end{bmatrix} \right\} \mid x, y, z \in \mathbb{R} \right\} \in \mathbb{R}^3$$

5 Scree Plot

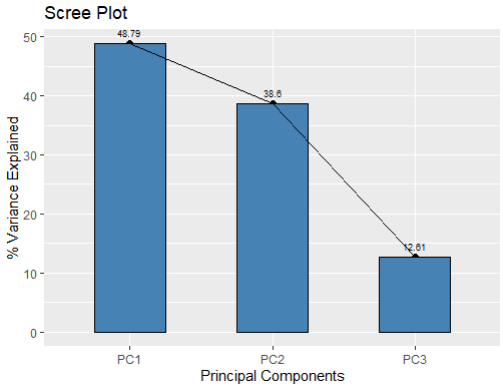
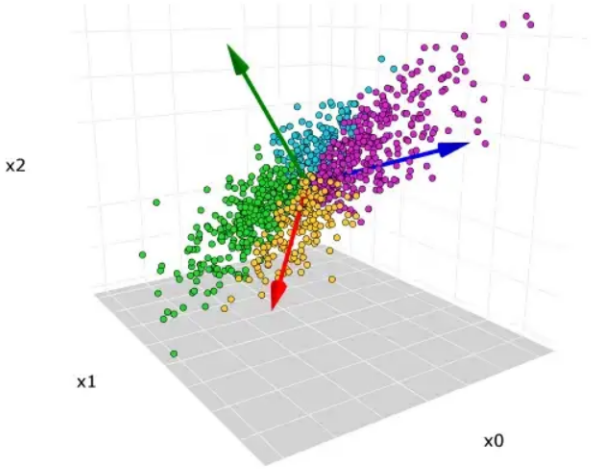


Figure 5: *a scree plot that summarizes the percentage of explained variance by each eigenvector*

5.1 What is a Scree Plot?

A scree plot is a useful tool when performing PCA. It summarizes how much variance is explained by our principal components, otherwise known as, the eigenvectors. After running the data in section 4.3 through a PCA analysis in R, I created this scree plot to summarize how much information is explained by each principal component. As we can see, PC1's direction contains the most variance at 49%, whereas, PC2's direction contains 39% ,the second most, and PC3's direction contains the least at 13%. If I were an analyst looking to build a model with these results I would most certainly want to try to include all of these components. This is because three features is not a lot when it comes to model building. Disregarding 12% of the variance when the number of features is so little can be disastrous and can lead to inaccurate results. Like stated in section two, PCA really shines when there is 5 or more features.

6 Conclusion



[7]

Figure 6: *centered 3-dimensional data that is close to a 2-dimensional subspace with associated eigenvectors*

In conclusion, principal component analysis can be defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some scaler projection of the data comes to lie on the first component, the second greatest on the second component, and so on.[8]

In other words, PCA finds the lines of best fit, rotates the eigenvectors and transforms the data, such that, the shadows, or images of the data, in the nth dimension, get projected down into lower dimensions. You can think of it as trying to figure out which direction you should view your data so that the distance between each data point is maximized.

It will generally be the case in PCA, that the number of principal components is equal to the number of eigenvalues. Once we find our eigenvectors we rotate them such that they become our new axes through which we view the variance maximized data. One eigenvector for every feature. The best part of this dimensionality reduction technique is that it will generally be the case that only the first few principal components will be referenced in our final analysis.

References

[1] Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space". Philosophical Magazine. 2 (11): 559–572. doi:10.1080/14786440109462720.

[2] Guide to principal component analysis. (n.d.). https://medium.com/analytics-vidhya/guide-to-principal-component-analysis-ab04a8a9c305

[3] Kumar, R. (2018, June 19). Understanding Principal Component Analysis - Rishav Kumar. *Medium*. URL:https://medium.com/@aptrishu/understanding-principle-component-analysis-e32be0253ef0 Retrieved. December 2022

[4] StatQuest with Josh Starmer. (2018, April 2). *StatQuest: Principal Component Analysis*(PCA), *Step-by-Step* [Video]. YouTube. URL: https://www.youtube.com/watch?v=FgakZw6K1QQ Retrieved. October 2022

[5] Wikipedia contributors. (2023, January 26). *Principal component analysis*. Wikipedia. URL:https://en.wikipedia.org/wiki/Principal_component_analysis Retrieved. December 2022

[6] *Matrix Theory and Linear Algebra : An open text by Peter Selinger, Based on the original text by Lyryx Learning and Ken Kuttler*, 1st ed. Selinger, P. (2018), Creative Commons License, URL:https://creativecommons.org/licenses/by/4.0/ URL:https://github.com/selinger/linear-algebra

[7] Cheng, C. (2022, September 9). Principal Component Analysis (PCA) Explained Visually with Zero Math. Medium. URL: https://towardsdatascience.com/principal-component-analysis-pca-explained-visually-with-zero-math-1cbf392b9e7d

[8] PCA - CS 357. (n.d.-b). URL: https://courses.grainger.illinois.edu/cs357/fa2022/notes/ref-18-pca.html