MATH 2101 Linear Algebra: Principal Component Analysis

Andrew O'Drain

December 4th 2022

Introduction

Principal component analysis is a statistical technique that is used to reduce the dimensionality of very 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". On its surface, principal component analysis seems like a complicated venture, however, after reviewing the literature, it seems it can be broken down into relatively bite-sized and understandable chunks. That is one goal I hope to attain in writing this paper. My most important, and main goal, is to demonstrate, with many examples from the linear algebra textbook, why principal component analysis is in-disposable as a dimensionality reduction technique in statistics.

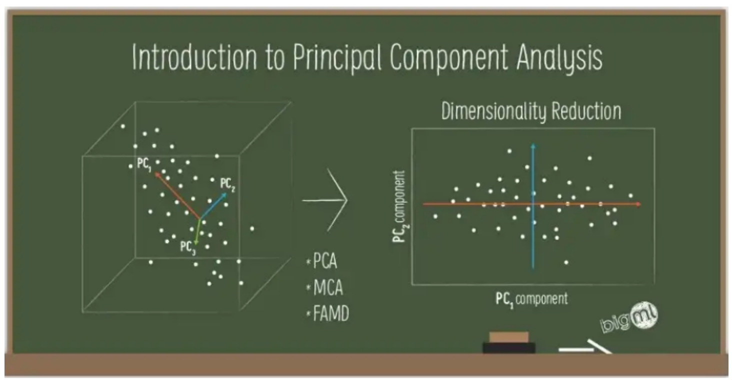

Figure 1: The centered 3-dimensional data with associated eigenvectors

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 in lay-men's terms, as best I can, (sometimes topics are so complicated that it's extremely difficult) what PCA is. But first, we need to understand what *features* are. Features are nothing more than another name for random variables in a data set. Features could be of the discrete data type, like the number of cars that pass on a certain 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 these features to extract meaningful insights from the data. You can think of the variance within features as the information of the data set. Sometimes, when there are a large number of features in a data set, the analyst wishes to measure how impactful different features are in explaining the variation in their dependent, or Y-variable. Ideally, the analyst wishes to keep one-hundred percent of the meaningful variation among and within 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. These are the new features that are linear combinations, or, in other words, mixtures of the initial features such that most of the information or variance is compressed into the first feature. For example, a 10 dimensional data set, a data set with 10 features, will yield 10 principal components, but PCA tries to put the maximum possible variance into the first variable, then the second and so on until we can formulate a graph that summarizes the percentage of variance explained by each eigenvector. This graph is called a scree plot, which we will learn more about later. The scree plot allows the analyst to discard features with low variance, and keep only those that contribute the most information. Since each principal component is a linear combination of the original the new features are a bit more difficult to interpret. However, this is OK. Sometimes the analyst needs to sacrifice some precision for information gain.

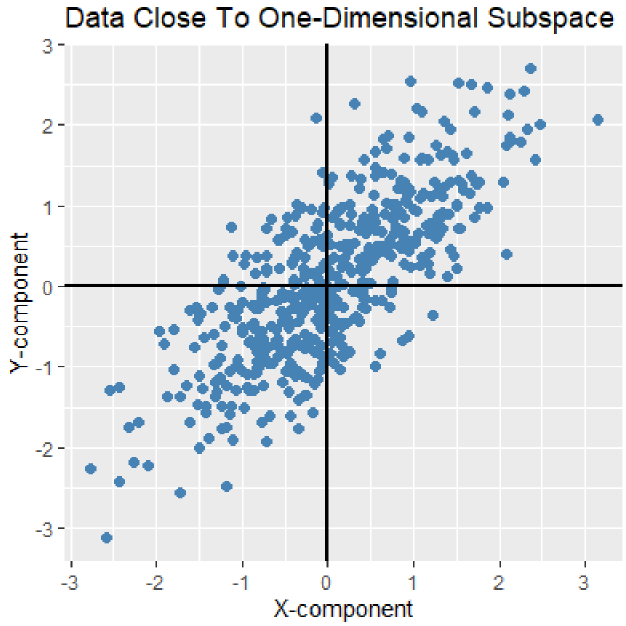

Figure 2: 2-dimensional data that is close to a 1-dimensional subspace W (a 1-dimensional line).

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 by projecting the image of the 2-dimensional subspace onto a 1-dimensional line? What would that look like? It would look like a one dimensional line such that the distances between each data point, the variance, is 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 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?

The Subspace Fitting Problem: Already Centered Data

Given that PCA appears to be overall complex, the mathematical procedures are straightforward and intuitive. The task of finding principal components is a bit easier if our data already passes through the origin. *Consider the following collection of points in* $\mathbb{R}^2$. (**Exercise 11.12.1**)

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

We are going to find the 1-dimensional subspace that best approximates this collection of points. We will also find the total squared distance of the 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 mean. Lets create a 2x10 matrix so we can perform the necessary calculations

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

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

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

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

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

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

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

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

As we can see we have two eigenvalues and two eigenvectors. If we pick the largest eigenvalue, $\lambda = 52$, and its corresponding eigenvector $\bar{v} = (-\frac{3}{2}, 1)$. This is the closest 1-dimensional subspace that is spanned by $\bar{v}$

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 sub-space, we will also find the closest 2-dimensional sub-space as well as the closest 3-dimensional sub-space. *Consider the following collection of points in* $\mathbb{R}^3$.(**Exercise 11.12.2**)

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

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

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

$$\begin{aligned} p_B(\lambda) &= -\lambda^3 - 395\lambda^2 + 24548\lambda - 417316 \\ &= -(\lambda - 323)(\lambda - 38)(\lambda - 34) \end{aligned}$$

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

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

Now that we have our eigenvalues in decreasing order with their associated eigenvectors, we can find the closest sub-spaces 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.

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

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

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

Affine Subspace Fitting: Data that does not pass through the origin

The above examples are an ideal scenario regarding fitting high dimensional data to lower dimensional sub-spaces. 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 coordinate where $\bar{y}$ 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 each other. It only shifts the data so it is mathematically centered about the origin.

To give you, the reader, a better intuition of what it means to include the centroid in our calculations, we will compute the centroid in the next example, 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* $\mathbb{R}^2$.(**Exercise 11.12.5**).

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

Compute Centroid & Center Data

$$\begin{aligned} B &= \frac{1}{10} \left(\begin{bmatrix} 4 \\ -3 \end{bmatrix} + \begin{bmatrix} 10 \\ -9 \end{bmatrix} + \begin{bmatrix} 4 \\ -7 \end{bmatrix} + \begin{bmatrix} -2 \\ 3 \end{bmatrix} + \begin{bmatrix} 10 \\ -3 \end{bmatrix} + \begin{bmatrix} 4 \\ -3 \end{bmatrix} + \begin{bmatrix} 5 \\ -3 \end{bmatrix} + \begin{bmatrix} 4 \\ 1 \end{bmatrix} + \begin{bmatrix} -2 \\ -3 \end{bmatrix} + \begin{bmatrix} 3 \\ -3 \end{bmatrix} \right) \\ &= \frac{1}{10} \begin{bmatrix} 40 \\ -30 \end{bmatrix} \\ &= \begin{bmatrix} 4 \\ -3 \end{bmatrix} \end{aligned}$$

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

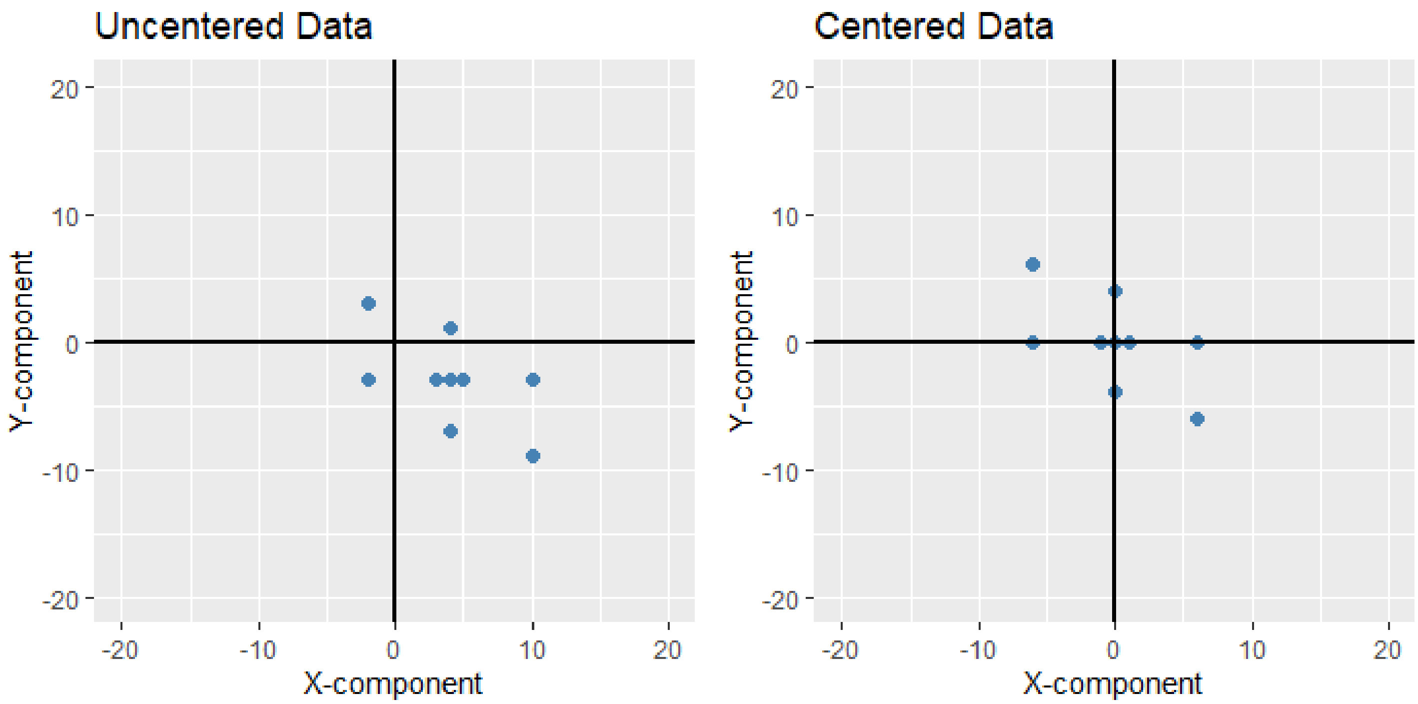

Figure 3: Figure 3: centered & un-centered data on a 2-dimensional subspace that is close to a 1-dimensional subspace

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

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

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

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

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

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

$$B + W = \{b + w \mid w \; \epsilon \; W\} = \left\{ \begin{bmatrix} 4 \\ -3 \end{bmatrix} + y \begin{bmatrix} \text{-}\frac{4}{3} \\ 1 \end{bmatrix} \mid y \; \epsilon \; \mathbb{R} \right\}$$

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 and the total squared distance is the λ_2.

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

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

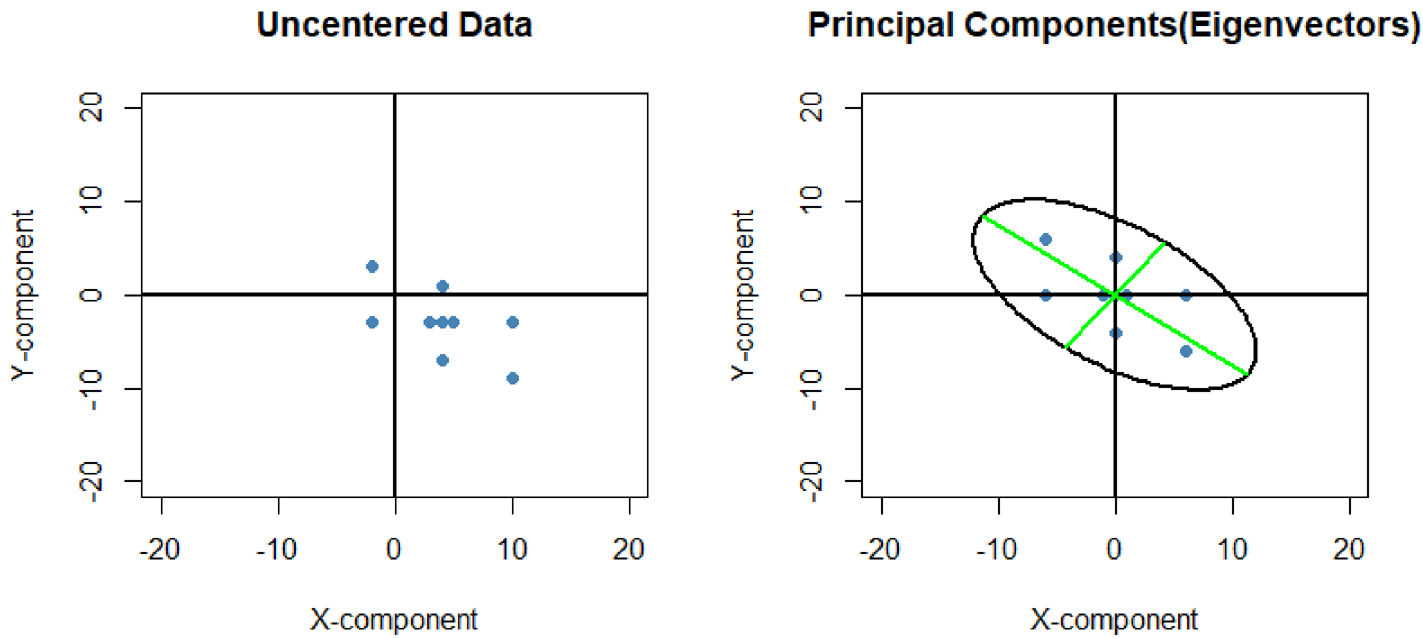

Figure 4: centered data & data with principal components overlaid. The longest green line is PC1 and the green line that is orthogonal to PC1 is PC2. As you can see, these eigenvectors span the direction of greatest variance

For our final example we will use three dimensions to tie all of our 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 our closest 1 and 2 dimensional affine sub-spaces that minimizes the square distances from our data to the subspace. Consider the following collection of points in $\mathbb{R}^3$(**Exercise 11.12.6**)

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

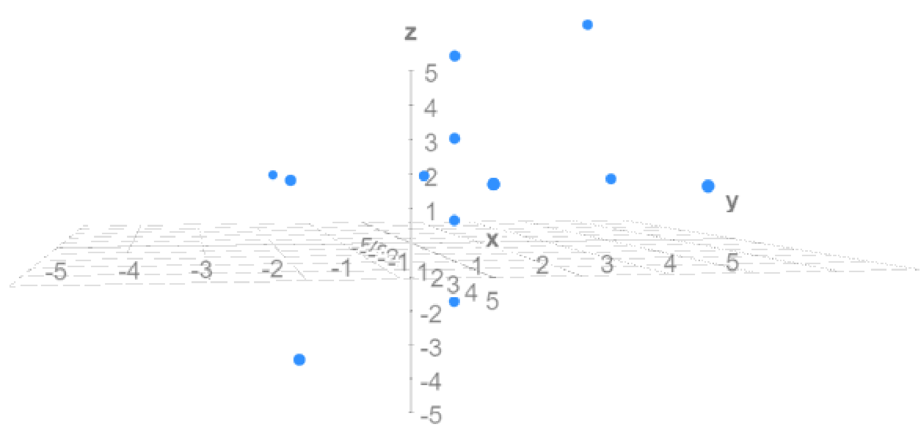

Figure 5: Arbitrary 3-dimensional data

Center Data

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

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

$$= \begin{bmatrix}3\\0\\2\end{bmatrix}$$

Subtract the Centroid from each Vector

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

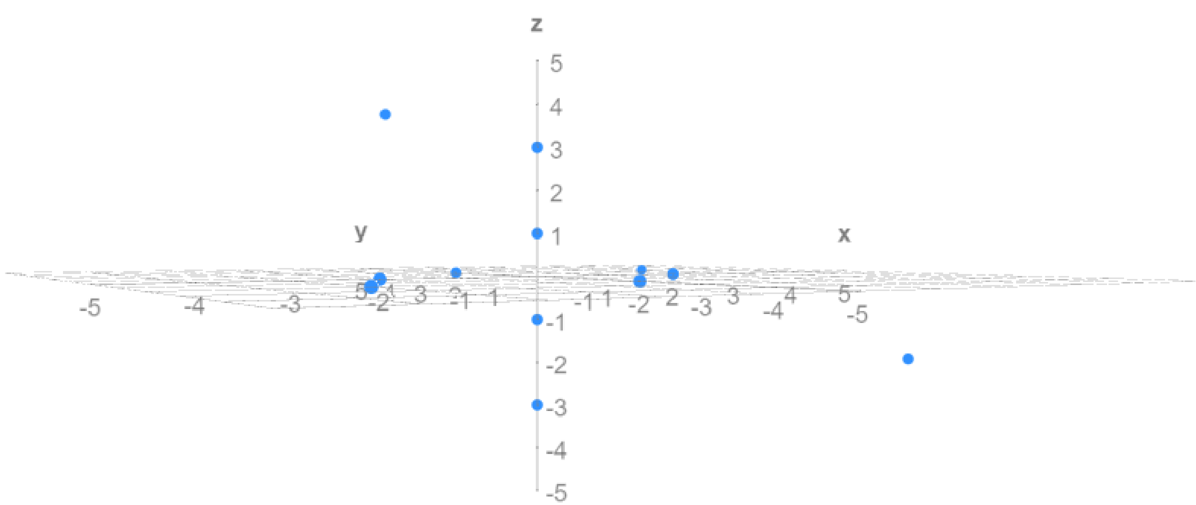

Figure 6: 3-dimensional data after subtracting the centroid & centering

Symmetric Matrix, Eigenvalues & Eigenvectors

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

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

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

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

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

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

Find Squared distances to $\mathbb{R}^n$

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

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

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

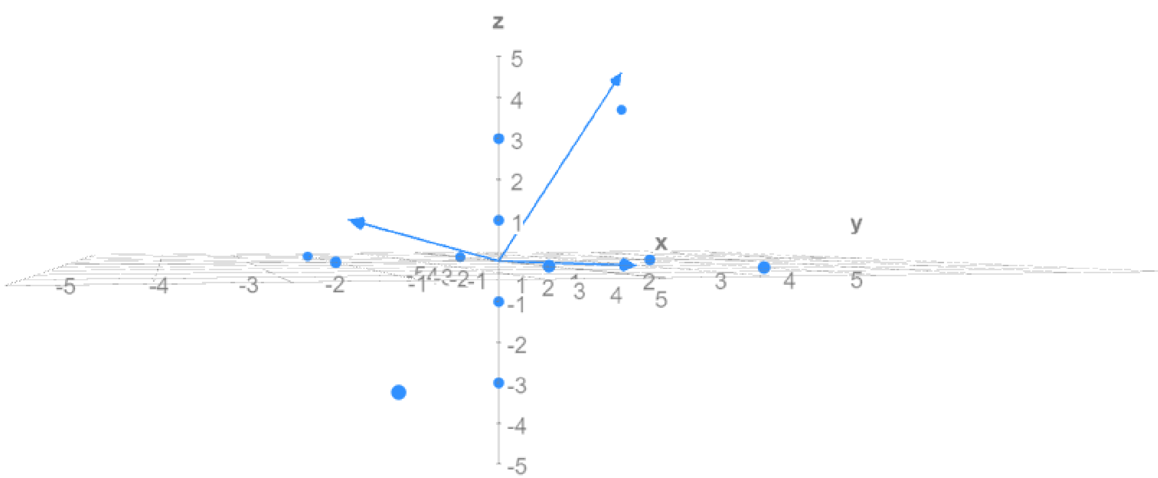

Figure 7: The centered 3-dimensional data with associated eigenvectors

Since we now have our eigenvectors we know in which direction the variance is maximized.

$$B + W = \{b + w \mid w \in W\} = \left\{ \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} + z \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \mid z \in \mathbb{R} \right\}$$

$$B + W = \{b + w \mid w \in W\} = \left\{ \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} + \left\{ z \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, z \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \right\} \mid z \in \mathbb{R} \right\}$$

To gain a more intuitive understanding of PCA you can think of it as trying to figure out which direction you should view your data from to extract the maximum amount of information. This is precisely what our eigenvalues tell us. Here we have three eigenvectors because we have three variables *x, y, z.* This will generally be the case in PCA. The number of principal components will equal the number of eigenvectors, which will equal the number of features of your data set which will equal the number of new axes on our scatter-plot. 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 it will generally be the case that only the first few principal components will be referenced in our final analysis. This also implies being able to visualize 100-dimensional data projected into 3-dimensions.

Scree Plot

When researching about principal component analysis I became aware that there is a specific type of plot that is used to summarize the the results of PCA analysis called a scree plot.

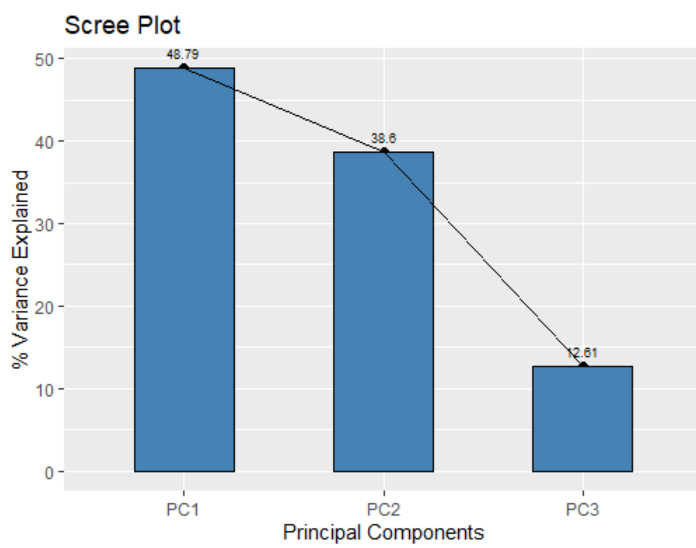

Figure 8: scree plot that summarizes the percentage of explained variance by each eigenvector

A scree plot is a useful tool when performing PCA. It summarizes how much variance is explained by our principal components, otherwise known as, eigenvectors. After running the PCA analysis through the programming language called 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.

In Summary

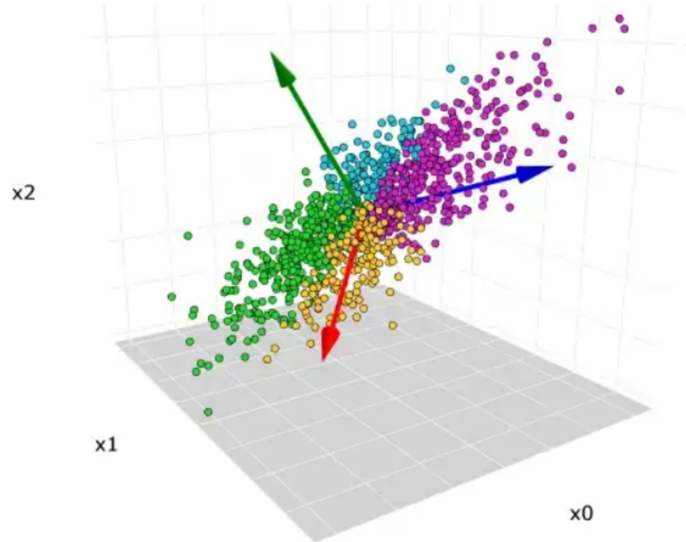

Figure 9: The centered 3-dimensional data with associated eigenvectors

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. In more simple terms, PCA finds the center of your data in n-dimensions, finds the lines of best fit; which are the eigenvalues, rotates and transforms your data such that the shadows, or images of your data from the nth dimension get projected onto your 3-dimensional, 2-dimensional, or 1-dimensional scatter-plot such that, the distances between each data point; the variance, is maximized. Hopefully you now have a better understanding of this marvelous dimensionality reduction technique. For a look at the parts of the calculations that were done by hand, check the following pages.

I found all of my research about PCA to be interesting. Before I began I had a summary understanding of what PCA was. I knew it was a dimensionality reduction technique, and that it involved eigenvectors, but I never firmly grasped exactly what PCA was doing. I especially found the scree plot and viewing variance as information interesting. The plot itself essentially gives you a quantified summary of how much insight one is able to extract from a data set. I also find it interesting that you can extrapolate this idea to the rest of statistics, and on its face, it seems to apply to every facet. Even though I have a more solid understanding of this topic, I am eager to learn if there are any other ways to perform the analysis. With all of this being said this semester has been rigorous as well as enlightening. Even though I will no longer be attending any formal classes on the topic, my goal is to continue teaching myself linear algebra indefinitely. Calculus used to be in the number one spot of my favorite branch of mathematics. I can now safely say that that spot goes to linear algebra now! Have a great holiday professor. I enjoyed the semester very much!

References

Medium, (2018, Jan 2) Understanding Principal Component Analysis.
https://medium.com/@aptrishu/understanding-principle-component-analysis-e32be0253ef0

Stat Quest. Principal Component Analysis Step-By-Step
https://www.youtube.com/watch?v=FgakZw6K1QQ

Wikipedia. Principal Component Analysis
https://en.wikipedia.org/wiki/Principal$_{c}omponent_{a}nalysis$

Calculations

11.12.1

$$A^TA = \begin{bmatrix} 1(1)-2(-2)+3(3)+2(2)-2(-2)+3(3)-3(-3)+0(0)+0(0) & 1(-1)-2(1)+3(-2)+2(-2)-2(1)+3(-1)-3(0)+0(3)+0(0)+0(-2) \\ -1(1)+1(-2)-2(3)-2(2)+1(-2)-1(3)+0(-3)+3(0)+0(0)-2(0) & -1(-1)+1(1)-2(-2)-2(-2)+1(1)-1(-1)+0(0)+3(3)+0(0)-2(-2) \end{bmatrix}$$

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

$$\det(B-\lambda_n I) = \begin{bmatrix} 40-\lambda_n & -18 \\ -18 & 25-\lambda_n \end{bmatrix} = (40-\lambda_n)(25-\lambda_n)-[-18(-18)] = \lambda^2-65\lambda+676 \qquad \lambda_1 = 52 \quad \lambda_2 = 13$$

$$= (x-52)(x-13)$$

$$B-\lambda_1 I = \begin{bmatrix} 40-52 & -18 \\ -18 & 25-52 \end{bmatrix} \qquad B-\lambda_2 I = \begin{bmatrix} 40-13 & -18 \\ -18 & 25-13 \end{bmatrix}$$

$$E_1(\lambda_1) = \begin{bmatrix} -12 & -18 \\ -18 & -27 \end{bmatrix} \quad \begin{bmatrix} 1 & 3/2 \\ 0 & 0 \end{bmatrix} \qquad E_2(\lambda_2) = \begin{bmatrix} 27 & -18 \\ -18 & 12 \end{bmatrix}$$

$-\frac{1}{12}R_1 \to R_1$

$$\begin{bmatrix} 1 & 3/2 \\ -18 & -27 \end{bmatrix} \quad \bar{V}_1 = \begin{bmatrix} -3/2 \\ 1 \end{bmatrix}$$

$18R_1+R_2 \to R_2$

$$\begin{bmatrix} x \\ y \end{bmatrix} = y\begin{bmatrix} -\frac{3}{2} \\ 1 \end{bmatrix}$$

$\frac{1}{27}R_1 \to R_1$

$$\begin{bmatrix} 1 & -2/3 \\ -18 & 12 \end{bmatrix} \quad \bar{V}_2 = \begin{bmatrix} 2/3 \\ 1 \end{bmatrix}$$

$18R_1+R_2 \to R_2$

$$\begin{bmatrix} 1 & -2/3 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} x \\ y \end{bmatrix} = y\begin{bmatrix} \frac{2}{3} \\ 1 \end{bmatrix}$$

Figure 10: 11.12.1

11.12.2

$$A^TA = \begin{bmatrix} 1(1)+6(6)+1(1)+2(2)+4(4)+1(1) & -2(6)-3(1) & -3(1)+6(13) \\ -2(6)-3(1) & 5(5)-2(-2)+2(2)+1(1)-3(3) & 5(-4)-2(13)+1(1) \\ -3(1)+13(6) & -4(5)+13(-2)+1(1) & 9(9)-3(-3)+1(1)-4(-4)+13(13)+1(1)+4(4) \end{bmatrix}$$

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

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

$$= 75(43-\lambda_n)75 - 45(-45)(59-\lambda_n) + (293-\lambda_n)(-15)(-15)$$

$$- \left[(59-\lambda_n)(43-\lambda_n)(293-\lambda_n) + (-15)(-45)(75) + (75)(-15)(-45)\right]$$

$$= -x^3 - 395x^2 + 24548x - 417316 \qquad \lambda_1 = 323 \quad \lambda_2 = 38 \quad \lambda_3 = 34$$

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

Figure 11: 11.12.2

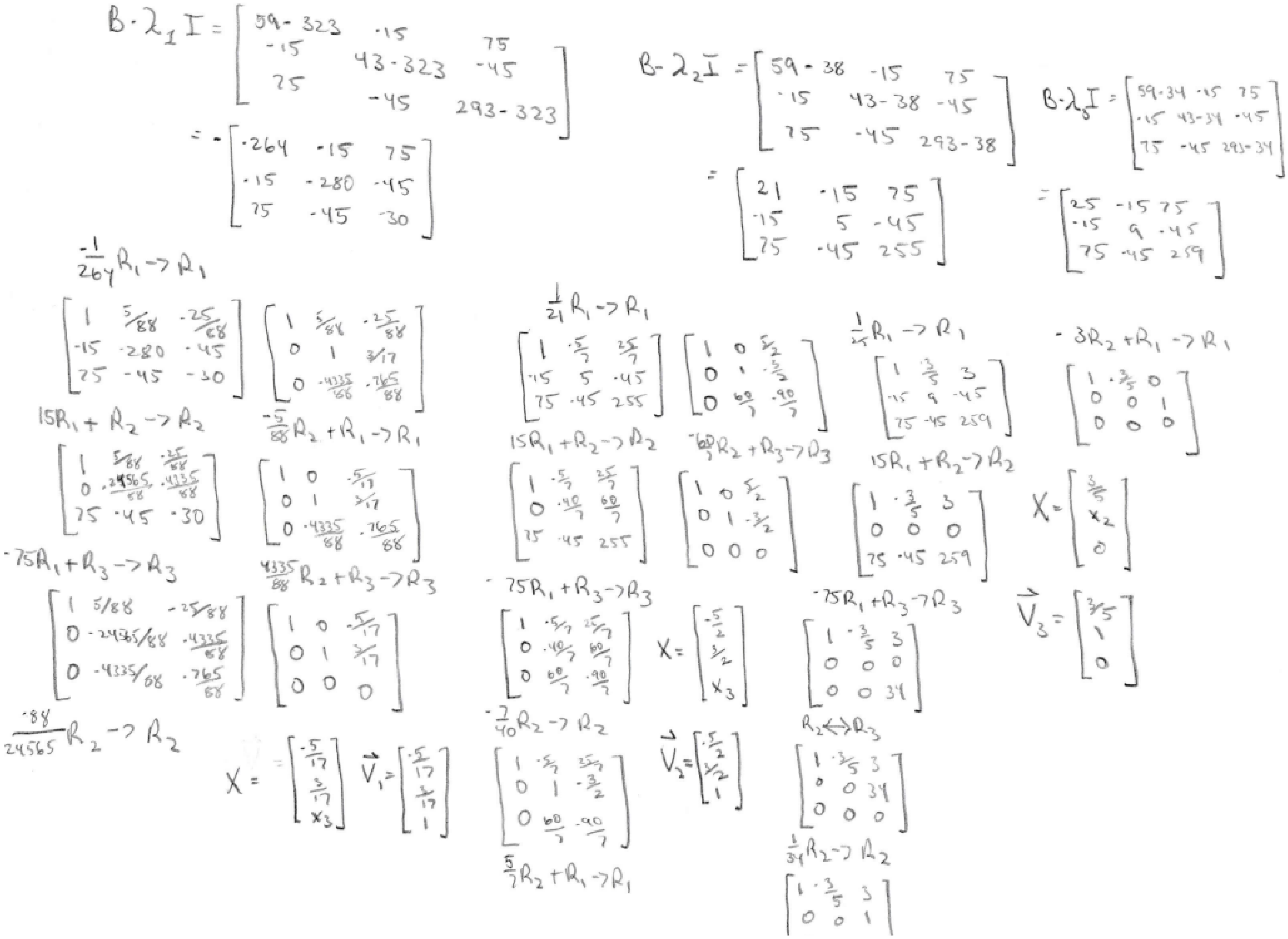

Figure 12: 11.12.2 cont.

11.12.3

$$A^TA = \begin{bmatrix} -1(-1)+3(3)+1(1) & -1(2)+1(-1) & -1(-1)+1(2) \\ 2(-1)+1(-1) & 1(1)+2(2)+3(3)+2(2)-1(-1) & 1(-1)+2(-1)+2(-2)-1(2) \\ -1(-1)+2(1) & -1(1)-1(2)-2(2)+2(-1) & 1(1)-1(-1)-3(-3)+3(3)-1(-1)-2(-2)+2(2) \end{bmatrix}$$

$$B = \begin{bmatrix} 11 & -3 & 3 \\ -3 & 19 & -9 \\ 3 & -9 & 29 \end{bmatrix}$$

$$\det(B - \lambda_n I) = \begin{bmatrix} 11-\lambda_n & -3 & 3 \\ -3 & 19-\lambda_n & -9 \\ 3 & -9 & 29-\lambda_n \end{bmatrix} \begin{matrix} 11-\lambda_n & -3 \\ -3 & 19-\lambda_n \\ 3 & -9 \end{matrix} = 3(19-\lambda_n)3 - 9(-9)(11-\lambda_n) + (29-\lambda_n)(-3)(-3)$$

$$-\left[(11-\lambda_n)(19-\lambda_n)(29-\lambda_n) - 3(-9)(3) + 3(-3)(-9)\right]$$

$$= -x^3 + 59x^2 - 980x + 4900$$

$$= -(\lambda-35)(\lambda-14)(\lambda-10) \quad \lambda_1 = 35 \quad \lambda_2 = 14 \quad \lambda_3 = 10$$

Figure 13: 11.12.3

11.12.3 cont

$$B-\lambda_1 I = \begin{bmatrix} 11-35 & -3 & 3 \\ -3 & 19-35 & -9 \\ 3 & -9 & 29-35 \end{bmatrix} = \begin{bmatrix} -24 & -3 & 3 \\ -3 & -16 & -9 \\ 3 & -9 & -6 \end{bmatrix}$$

$-\frac{1}{24}R_1 \to R_1$
$$\begin{bmatrix} 1 & 1/8 & -\frac{1}{8} \\ -3 & -16 & -9 \\ 3 & -9 & -6 \end{bmatrix}$$

$-3R_1 + R_2 \to R_2$
$$\begin{bmatrix} 1 & 1/8 & -1/8 \\ 0 & -125/8 & -75/8 \\ 3 & -9 & -6 \end{bmatrix}$$

$-3R_1 + R_3 \to R_3$
$$\begin{bmatrix} 1 & 1/8 & -\frac{1}{8} \\ 0 & -125/8 & -\frac{75}{8} \\ 0 & -\frac{75}{8} & -\frac{45}{8} \end{bmatrix}$$

$-\frac{8}{125}R_2 \to R_2$ $\quad -\frac{1}{8}R_2 + R_1 \to R_1$ $\quad \frac{75}{8}R_2 + R_3 \to R_3$
$$\begin{bmatrix} 1 & 1/8 & -\frac{1}{8} \\ 0 & 1 & 3/5 \\ 0 & -\frac{75}{8} & -\frac{45}{8} \end{bmatrix} \begin{bmatrix} 1 & 0 & -1/5 \\ 0 & 1 & 3/5 \\ 0 & -\frac{75}{8} & -\frac{45}{8} \end{bmatrix} \begin{bmatrix} 1 & 0 & -1/5 \\ 0 & 1 & 3/5 \\ 0 & 0 & 0 \end{bmatrix}$$

$$X = \begin{bmatrix} 1/5 \\ -3/5 \\ x_3 \end{bmatrix} \quad \vec{V}_1 = \begin{bmatrix} 1/5 \\ -3/5 \\ 1 \end{bmatrix}$$

$$B-\lambda_2 I = \begin{bmatrix} 11-14 & -3 & 3 \\ -3 & 19-14 & -9 \\ 3 & -9 & 29-14 \end{bmatrix} = \begin{bmatrix} -3 & -3 & 3 \\ -3 & 5 & -9 \\ 3 & -9 & 15 \end{bmatrix}$$

$-\frac{1}{3}R_1 \to R_1$
$$\begin{bmatrix} 1 & 1 & -1 \\ -3 & 5 & -9 \\ 3 & -9 & 15 \end{bmatrix}$$

$3R_1 + R_2 \to R_2$
$$\begin{bmatrix} 1 & 1 & -1 \\ 0 & 8 & -12 \\ 3 & -9 & 15 \end{bmatrix}$$

$-3R_1 + R_3 \to R_3$
$$\begin{bmatrix} 1 & 1 & -1 \\ 0 & 8 & -12 \\ 0 & -12 & 18 \end{bmatrix}$$

$\frac{1}{8}R_2 \to R_2$
$$\begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & -3/2 \\ 0 & -12 & 18 \end{bmatrix} \quad X = \begin{bmatrix} -\frac{1}{2} \\ 3/2 \\ x_3 \end{bmatrix} \quad \vec{V}_2 = \begin{bmatrix} -1/2 \\ 3/2 \\ 1 \end{bmatrix}$$

$-R_2 + R_1 \to R_1$ $\quad 12R_2 + R_3 \to R_3$
$$\begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & -3/2 \\ 0 & -12 & 18 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

$$B-\lambda_3 I = \begin{bmatrix} 11-10 & -3 & 3 \\ -3 & 19-10 & -9 \\ 3 & -9 & 29-10 \end{bmatrix} = \begin{bmatrix} 1 & -3 & 3 \\ -3 & 9 & -9 \\ 3 & -9 & 19 \end{bmatrix}$$

$3R_1 + R_2 \to R_2$
$$\begin{bmatrix} 1 & -3 & 3 \\ 0 & 0 & 0 \\ 3 & -9 & 19 \end{bmatrix}$$

$-3R_1 + R_3 \to R_3$
$$\begin{bmatrix} 1 & -3 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 10 \end{bmatrix}$$

$R_2 \leftrightarrow R_3 \to R_2$
$$\begin{bmatrix} 1 & -3 & 3 \\ 0 & 0 & 10 \\ 0 & 0 & 0 \end{bmatrix}$$

$\frac{1}{10}R_2 \to R_2$
$$\begin{bmatrix} 1 & -3 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad X = \begin{bmatrix} 3 \\ x_2 \\ 0 \end{bmatrix}$$

$-3R_2 + R_1 \to R_1$
$$\begin{bmatrix} 1 & -3 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \vec{V}_3 = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$$

Figure 14: 11.12.3 cont.

11.12.4 Centroid

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

$$= \frac{1}{10}\begin{bmatrix} -4 \\ -22 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{2}{5} \\ \frac{11}{5} \end{bmatrix}$$

Figure 15: 11.12.4

11.12.5

$$A^TA = \begin{bmatrix} -6(-6)+6(6)-6(-6)-1(-1)+6(6) & -6(6)+6(-6) \\ 6(-6)-6(6)-1(-4) & 6(6)+4(4)-6(-6)-4(-4) \end{bmatrix}$$

$$B = \begin{bmatrix} 146 & -72 \\ -72 & 104 \end{bmatrix}$$

$$\det(B - \lambda_n I) = \begin{bmatrix} 146-\lambda_n & -72 \\ -72 & 104-\lambda_n \end{bmatrix} = (146-\lambda_n)(104-\lambda_n) - [-72(-72)] = \lambda^2 - 250x + 10000$$

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

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

$$B - \lambda_1 I = \begin{bmatrix} -54 & -72 \\ -72 & -96 \end{bmatrix}$$

$$-\tfrac{1}{54}R_1 \to R_1 \qquad \begin{bmatrix} 1 & 4/3 \\ -72 & -96 \end{bmatrix}$$

$$72R_1 + R_2 \to R_2 \qquad \begin{bmatrix} 1 & 4/3 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = x_2 \begin{bmatrix} -\frac{4}{3} \\ 1 \end{bmatrix} \quad \therefore \bar{v}_1 = \begin{bmatrix} -\frac{4}{3} \\ 1 \end{bmatrix}$$

$$B - \lambda_2 I = \begin{bmatrix} 96 & -72 \\ -72 & 54 \end{bmatrix}$$

$$\tfrac{1}{96}R_1 \to R_1 \qquad \begin{bmatrix} 1 & -3/4 \\ -72 & 54 \end{bmatrix}$$

$$72R_1 + R_2 \to R_2 \qquad \begin{bmatrix} 1 & -3/4 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = x_2 \begin{bmatrix} \frac{3}{4} \\ 1 \end{bmatrix} \quad \therefore \bar{v}_2 = \begin{bmatrix} \frac{3}{4} \\ 1 \end{bmatrix}$$

Figure 16: 11.12.5

$$A^TA = \begin{bmatrix} -3(-3)+3(3)-1(-1)+2(2)+1(1)+2(2)-2(-2)-2(-2) & -3(-2)+3(2)-1(2)+1(-2) & -1(4)+1(-4) \\ -2(-3)+2(3)-2(1)-2(1) & -2(-2)+2(2)+2(2)-2(-2)-2(-2) & 2(4)-2(-4) \\ 4(-1)-4(1) & 4(2)-4(-2) & 4(4)+3(3)-4(-4)-3(-3)-1(-1)+1(1) \end{bmatrix}$$

$$B = \begin{bmatrix} 36 & 8 & -8 \\ 8 & 24 & 16 \\ -8 & 16 & 52 \end{bmatrix}$$

$$\det(B-\lambda_n I) = \begin{bmatrix} 36-\lambda_n & 8 & -8 \\ 8 & 24-\lambda_n & 16 \\ -8 & 16 & 52-\lambda_n \end{bmatrix} \begin{matrix} 36-\lambda_n & 8 \\ 8 & 24-\lambda_n \\ -8 & 16 \end{matrix} = (36-\lambda_n)(24-\lambda_n)(52-\lambda_n)+(8)(16)(-8)+(-8)(8)(16) - \left[(-8)(24-\lambda_n)(-8)+(16)(16)(36-\lambda_n)+(52-\lambda_n)(8)(8)\right]$$

$$= -\lambda^3+112\lambda^2-3600\lambda+28800 \qquad \lambda_1=12 \quad \lambda_2=40 \quad \lambda_3=60$$

$$= -(\lambda-12)(\lambda^2-100\lambda+2400)$$

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

$$(B-\lambda_1 I) = \begin{bmatrix} 36-12 & 8 & -8 \\ 8 & 24-12 & 16 \\ -8 & 16 & 52-12 \end{bmatrix}$$

$$= \begin{bmatrix} 24 & 8 & -8 \\ 8 & 12 & 16 \\ -8 & 16 & 40 \end{bmatrix} \xrightarrow{\frac{1}{24}R_1 \to R_1} \begin{bmatrix} 1 & 1/3 & -1/3 \\ 8 & 12 & 16 \\ -8 & 16 & 40 \end{bmatrix} \xrightarrow{-8R_1+R_2\to R_2} \begin{bmatrix} 1 & 1/3 & -1/3 \\ 0 & 28/3 & 56/3 \\ -8 & 16 & 40 \end{bmatrix} \xrightarrow{8R_1+R_3\to R_3} \begin{bmatrix} 1 & 1/3 & -1/3 \\ 0 & 28/3 & 56/3 \\ 0 & 56/3 & 112/3 \end{bmatrix}$$

$$\xrightarrow{\frac{3}{28}R_2\to R_2} \begin{bmatrix} 1 & 1/3 & -1/3 \\ 0 & 1 & 2 \\ 0 & 56/3 & 112/3 \end{bmatrix} \xrightarrow{-\frac{1}{3}R_2+R_1\to R_1} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 56/3 & 112/3 \end{bmatrix} \xrightarrow{-\frac{56}{3}R_2+R_3\to R_3} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = z\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = E_{\lambda_1}$$

$$\text{Basis}(E_{\lambda_1}) = \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}$$

$$\underline{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Figure 17: 11.12.6

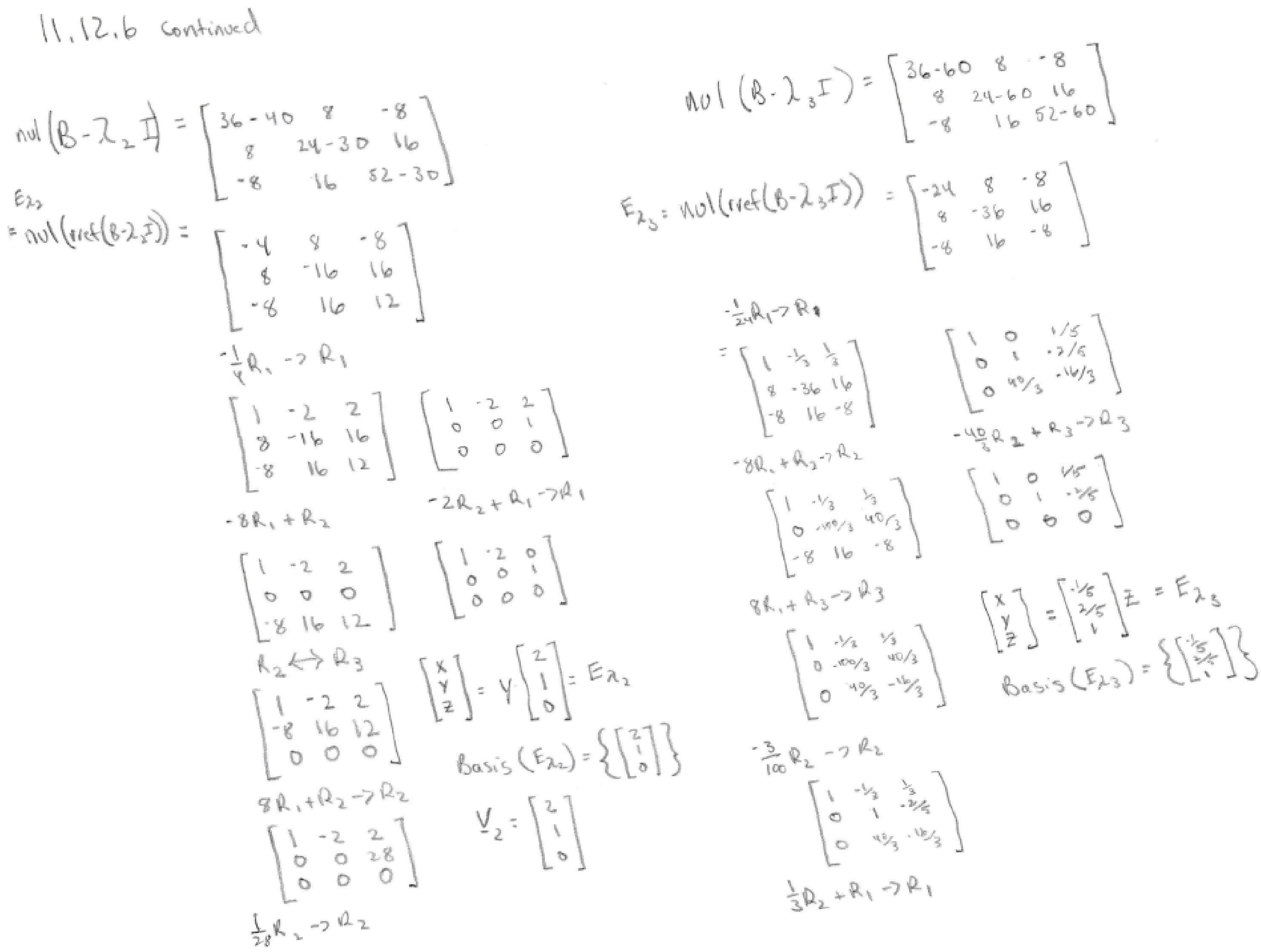

Figure 18: 11.12.6 cont