Eigenvalues, Eigenvectors, and an Intro to PCA
We’ve talked so far about re-writing our data using a new set of variables, or a new basis.

How to choose this new basis??
One (very popular) method: start by choosing the basis vectors as the directions in which the variance of the data is maximal.
One (very popular) method: start by choosing the basis vectors as the directions in which the variance of the data is maximal.
Then, choose subsequent directions that are orthogonal to the first and have the \emph{next} largest variance.
As it turns out, these directions are given by the eigenvectors of either:

- The Covariance Matrix (data is only centered)
- The Correlation Matrix (data is completely standardized)

Great. What the heck is an eigenvector?
Chapter ?
Eigenvalues and Eigenvectors
When we multiply a matrix by a vector, what do we get?

\[
\begin{pmatrix}
A
\end{pmatrix}
\begin{pmatrix}
x
\end{pmatrix}
= ?
= \begin{pmatrix}
b
\end{pmatrix}
\]

Another vector.

When the matrix $A$ is square ($n \times n$) then $x$ and $b$ are both vectors in $\mathbb{R}^n$. We can draw (or imagine) them in the same space.
In general, multiplying a vector by a matrix changes both its magnitude and direction.

\[
A = \begin{pmatrix} -1 & 1 \\ 6 & 0 \end{pmatrix} \quad x = \begin{pmatrix} 1 \\ 5 \end{pmatrix} \quad Ax = \begin{pmatrix} 4 \\ 6 \end{pmatrix}
\]
Eigenvectors

However, a matrix may act on certain vectors by changing only their magnitude, not their direction (or possibly reversing it)

\[
A = \begin{pmatrix} -1 & 1 \\ 6 & 0 \end{pmatrix} \quad x = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad Ax = \begin{pmatrix} 2 \\ 6 \end{pmatrix}
\]

The matrix \( A \) acts like a scalar!
For a square matrix, $A$, a non-zero vector $x$ is called an **eigenvector of $A$** if multiplying $x$ by $A$ results in a scalar multiple of $x$.

$$Ax = \lambda x$$

The scalar $\lambda$ is called the **eigenvalue** associated with the eigenvector.
\[
A = \begin{pmatrix} -1 & 1 \\ 6 & 0 \end{pmatrix} \quad x = \begin{pmatrix} 1 \\ 3 \end{pmatrix}
\]
\[
Ax = \begin{pmatrix} 2 \\ 6 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 3 \end{pmatrix} = 2x
\]

\[\implies \lambda = 2 \text{ is the eigenvalue}\]
Example 2

Show that $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ and find the corresponding eigenvalue.

\[
\mathbf{A} = \begin{pmatrix} -1 & 1 \\ 6 & 0 \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}
\]

\[
\mathbf{A}\mathbf{x} = \begin{pmatrix} -3 \\ 6 \end{pmatrix} = -3 \begin{pmatrix} 1 \\ -2 \end{pmatrix} = -3\mathbf{x}
\]

$\implies \lambda = -3$ is the eigenvalue corresponding to the eigenvector $\mathbf{x}$
Example 2

\[ A = \begin{pmatrix} -1 & 1 \\ 6 & 0 \end{pmatrix} \quad x = \begin{pmatrix} 1 \\ -2 \end{pmatrix} \quad Ax = -3x \]
Let’s Practice

1. Show that \( \mathbf{v} \) is an eigenvector of \( \mathbf{A} \) and find the corresponding eigenvalue:

\[
\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} 3 \\ -3 \end{pmatrix}
\]

\[
\mathbf{A} = \begin{pmatrix} -1 & 1 \\ 6 & 0 \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}
\]

2. Can a rectangular matrix have eigenvalues/eigenvectors?
Fact: Any scalar multiple of an eigenvector of $A$ is also an eigenvector of $A$ with the same eigenvalue:

$$A = \begin{pmatrix} -1 & 1 \\ 6 & 0 \end{pmatrix} \quad x = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad \lambda = 2$$

Try

$$v = \begin{pmatrix} 2 \\ 6 \end{pmatrix} \text{ or } u = \begin{pmatrix} -1 \\ -3 \end{pmatrix} \text{ or } z = \begin{pmatrix} 3 \\ 9 \end{pmatrix}$$

In general (#proof): Let $Ax = \lambda x$. If $c$ is some constant, then:

$$A(cx) = c(Ax) = c(\lambda x) = \lambda(cx)$$
For a given matrix, there are infinitely many eigenvectors associated with any one eigenvalue.

Any scalar multiple (positive or negative) can be used.

The collection is referred to as the eigenspace associated with the eigenvalue.

In the previous example, the eigenspace of \( \lambda = 2 \) was the span

\[
\text{span}\left\{ x = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \right\}.
\]

With this in mind what should you expect from software?
What if $\lambda = 0$ is an eigenvalue for some matrix $A$?

$$Ax = 0x = 0,$$  
where $x \neq 0$ is an eigenvector.

This means that some linear combination of the columns of $A$ equals zero!

$\implies$ Columns of $A$ are linearly dependent.

$\implies$ $A$ is not full rank.

$\implies$ Perfect Multicollinearity.
For a square \((n \times n)\) matrix \(A\), eigenvalues and eigenvectors come in pairs. Normally, the pairs are ordered by magnitude of eigenvalue.

\[
|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|
\]

\(\lambda_1\) is the largest eigenvalue

Eigenvector \(v_i\) corresponds to eigenvalue \(\lambda_i\)
Let's Practice

1. For the following matrix, determine the eigenvalue associated with the given eigenvector.

\[
A = \begin{pmatrix}
1 & -1 & 0 \\
0 & 1 & -1 \\
1 & 0 & -1 \\
\end{pmatrix} \quad v = \begin{pmatrix}
1 \\
1 \\
1 \\
\end{pmatrix}
\]

From this eigenvalue, what can you conclude about the matrix \( A \)?

2. The matrix \( M \) has eigenvectors \( u \) and \( v \). What is \( \lambda_1 \), the first eigenvalue for the matrix \( M \)?

\[
M = \begin{pmatrix}
-1 & 1 \\
6 & 0 \\
\end{pmatrix} \quad u = \begin{pmatrix}
1 \\
3 \\
\end{pmatrix} \quad v = \begin{pmatrix}
1 \\
-2 \\
\end{pmatrix}
\]

3. For the previous problem, is the specific eigenvector (\( u \) or \( v \)) the only eigenvector associated with \( \lambda_1 \)?
What happens if I put all the eigenvectors of $A$ into a matrix as columns:

$$V = [v_1 \mid v_2 \mid v_3 \mid \ldots \mid v_n]$$

and then compute $AV$?

$$AV = A[v_1 \mid v_2 \mid v_3 \mid \ldots \mid v_n]$$
$$= [Av_1 \mid Av_2 \mid Av_3 \mid \ldots \mid Av_n]$$
$$= [\lambda_1 v_1 \mid \lambda_2 v_2 \mid \lambda_3 v_3 \mid \ldots \mid \lambda_n v_n]$$

The effect is that the columns are all scaled by the corresponding eigenvalue.
Diagonalization

\[ AV = [\lambda_1 v_1 | \lambda_2 v_2 | \lambda_3 v_3 | \ldots | \lambda_n v_n] \]

The same thing would happen if we multiplied \( V \) by a diagonal matrix of eigenvalues on the right:

\[
VD = [v_1 | v_2 | v_3 | \ldots | v_n]
\begin{pmatrix}
\lambda_1 & 0 & 0 & \ldots & 0 \\
0 & \lambda_2 & 0 & \ldots & 0 \\
0 & 0 & \ddots & \ddots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \lambda_n
\end{pmatrix}
\]

\[
= [\lambda_1 v_1 | \lambda_2 v_2 | \lambda_3 v_3 | \ldots | \lambda_n v_n]
\]

\[ \implies AV = VD \]
\[ AV = VD \]

If the matrix \( A \) has \( n \) linearly independent eigenvectors, then the matrix \( V \) has an inverse.

\[ \implies V^{-1}AV = D \]

This is called the \textit{diagonalization} of \( A \).

It is only possible when \( A \) has \( n \) linearly independent eigenvectors (so that the matrix \( V \) has an inverse).
Let’s Practice

For the matrix

$$A = \begin{pmatrix} 0 & 4 \\ -1 & 5 \end{pmatrix},$$

a. The pairs

$$\lambda_1 = 4, \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \lambda_2 = 1, \mathbf{v}_2 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

are eigenpairs of $A$. For each pair, provide another eigenvector associated with that eigenvalue.

b. Based on the eigenvectors listed in part a, can the matrix $A$ be diagonalized? Why or why not? If diagonalization is possible, explain how it would be done.
Symmetric matrices (like the covariance matrix, the correlation matrix, and $X^TX$) have some very nice properties:

1. They are *always* diagonalizable. (Always have $n$ linearly independent eigenvectors.)
2. Their eigenvectors are all mutually orthogonal.
3. Thus, if you normalize the eigenvectors (software does this automatically) they will form an orthogonal matrix, $V$.

\[ AV = VD \]
\[ V^TAV = D \]
Eigenvectors of Symmetric Matrices
Introduction to Principal Components Analysis (PCA)
Eigenvalues and Eigenvectors
Covariance/Correlation matrices are symmetric. 

⇒ Eigenvectors are orthogonal.

Eigenvectors are ordered by the magnitude of eigenvalues:

\[ |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_p| \]

\[ \{v_1, v_2, \ldots, v_n\} \]
We’ll get into more detail later, but

- For the covariance matrix, we want to think of our data as *centered* to begin with.
- For correlation matrix, we want to think of our data as *standardized* to begin with. (i.e. centered *and* divided by standard deviation).
The first eigenvector of a covariance/correlation matrix points in the direction of maximum variance in the data. This eigenvector is the **first principal component**.

![Plot of height vs. weight with data points scattered around the origin.](image-url)
The first eigenvector of a covariance matrix points in the direction of maximum variance in the data. This eigenvector is the first principal component.
The first principal component is not the same direction as the regression line. It minimizes orthogonal distances, not vertical distances.
Introduction to Principal Components Analysis (PCA)

Not a Regression Line

It may be close in some cases, but it’s not the same.
The second eigenvector of a covariance matrix points in the direction, orthogonal to the first, that has the maximum variance.
Principal components provide us with a new orthogonal basis where the new coordinates of the data points are uncorrelated:
Each principal component is a linear combination of the original variables.

\[ \mathbf{v}_1 = \begin{pmatrix} 0.7 \\ 0.7 \end{pmatrix} = 0.7\mathbf{h} + 0.7\mathbf{w} \]

\[ \mathbf{v}_2 = \begin{pmatrix} -0.7 \\ 0.7 \end{pmatrix} = -0.7\mathbf{h} + 0.7\mathbf{w} \]

These coefficients are called **loadings** (factor loadings).
Likewise, we can think of our original variables as linear combinations of the principal components with coordinates in the new basis:

\[ h = \begin{pmatrix} 0.7 \\ -0.7 \end{pmatrix} = 0.7v_1 - 0.7v_2 \]

\[ w = \begin{pmatrix} 0.7 \\ 0.7 \end{pmatrix} = 0.7v_1 + 0.7v_2 \]
The Biplot

- Uncorrelated data points and variable vectors plotted on the same set of axis!
- Points in top right have largest weight values
- Points in top left have smallest height values
Introduction to Principal Components Analysis (PCA)
The variable loadings give us a formula for finding the coordinates of our data in the new basis.

\[
v_1 = \begin{pmatrix} 0.7 \\ 0.7 \end{pmatrix} \implies PC_1 = 0.7 \text{height} + 0.7 \text{weight}
\]

\[
v_2 = \begin{pmatrix} -0.7 \\ 0.7 \end{pmatrix} \implies PC_2 = -0.7 \text{height} + 0.7 \text{weight}
\]

From these loadings, we can get a sense of how important each original variable is to our new variables (the components).
Let’s Practice

1. For the following data plot, take your best guess and draw the directions of the first and second principal components.

2. Is there more than one answer to this question?
Suppose your data contained the variables \textit{VO2\_max}, \textit{mile pace}, and \textit{weight} in that order. The first principal component for this data is the eigenvector of the covariance matrix.

\[
\begin{pmatrix}
0.92 \\
0.82 \\
-0.51
\end{pmatrix}
\]

What variable is most important/influential to the first principal component?