Lectures on Multivariable Mathematics: Principal Component Analysis and Singular Value Decomposition

Dr. Chuck Rocca
roccac@wcsu.edu

Table of Contents

1. Objectives
2. Singular Value Decomposition
3. Means and Variance
4. Principal Component Analysis

Objectives

After this lesson you should be able to:

1. define the singular values of a matrix,

Objectives

After this lesson you should be able to:

1. define the singular values of a matrix,
2. use the singular values and corresponding eigenvectors to decompose a matrix,

Objectives

After this lesson you should be able to:

1. define the singular values of a matrix,
2. use the singular values and corresponding eigenvectors to decompose a matrix,
3. replace the singular value decomposition with the reduced singular value decomposition,

Objectives

After this lesson you should be able to:

1. define the singular values of a matrix,
2. use the singular values and corresponding eigenvectors to decompose a matrix,
3. replace the singular value decomposition with the reduced singular value decomposition,
4. construct the sample mean vector and the mean-deviation form for a set of data,

Objectives

After this lesson you should be able to:

1. define the singular values of a matrix,
2. use the singular values and corresponding eigenvectors to decompose a matrix,
3. replace the singular value decomposition with the reduced singular value decomposition,
4. construct the sample mean vector and the mean-deviation form for a set of data,
5. calculate the covariance matrix for a set of data,

Objectives

After this lesson you should be able to:

1. define the singular values of a matrix,
2. use the singular values and corresponding eigenvectors to decompose a matrix,
3. replace the singular value decomposition with the reduced singular value decomposition,
4. construct the sample mean vector and the mean-deviation form for a set of data,
5. calculate the covariance matrix for a set of data,
6. use a covariance matrix to determine the principal component of a set of data, and

Objectives

After this lesson you should be able to:

1. define the singular values of a matrix,
2. use the singular values and corresponding eigenvectors to decompose a matrix,
3. replace the singular value decomposition with the reduced singular value decomposition,
4. construct the sample mean vector and the mean-deviation form for a set of data,
5. calculate the covariance matrix for a set of data,
6. use a covariance matrix to determine the principal component of a set of data, and
7. explain why it is called the principal component.

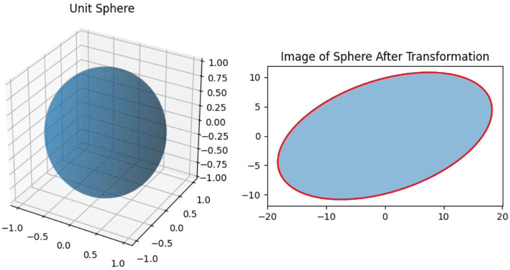
Table of Contents

1. Objectives
2. Singular Value Decomposition
3. Means and Variance
4. Principal Component Analysis

Motivating Example

A 2×3 matrix A is a transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$, for example:

$$\begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$$



Our goal is to find the major and minor axes of the ellipse.

Minimums and Maximums for Non-Square Transformations

The major and minor axes are the maximum and minimum length non-zero vectors $\vec{x}$ associated with multiplication by A:

$$||A\vec{x}||^2 = (A\vec{x})^T(A\vec{x}) = \left(\vec{x}^T A^T\right) A\vec{x} = \vec{x}^T \left(A^T A\right) \vec{x}$$

From before, the maximum and minimums associated with the symmetric matrix

$$B = A^T A = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

are given by the eigenvalues and vectors of B.

Minimums and Maximums for Non-Square Transformations

In the case of

$$B = A^T A = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

the eigenvalues are $\lambda_1 = 360$, $\lambda_2 = 90$, and $\lambda_3 = 0$ with corresponding vectors

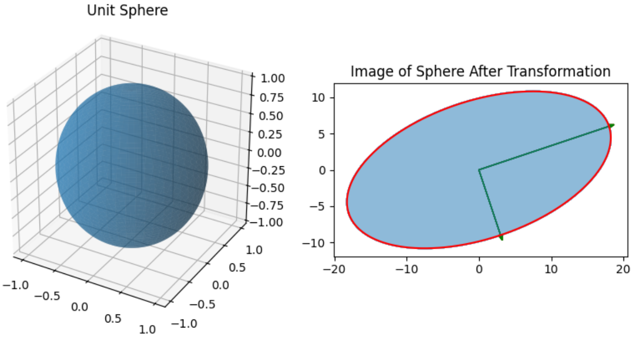
$$\vec{v}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \ \vec{v}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \text{ and } \vec{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

and multiplying by A we get

$$A\vec{v}_1 = \begin{bmatrix} 18 \\ 6 \end{bmatrix}, \ A\vec{v}_2 = \begin{bmatrix} 3 \\ -9 \end{bmatrix}, \text{ and } A\vec{v}_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Motivating Example Revisited

The major axis is in the direction of $A\overrightarrow{v}_1 = \langle 18, 6 \rangle$ and the minor is $A\overrightarrow{v}_2 = \langle 3, -9 \rangle$:



Singular Values of a Matrix

Definition (Singular Values of a Matrix)

The singular values of a matrix A are the square roots of the eigenvalues of $A^T A$. We assume that the singular values/eigenvalues are arranged in decreasing order and note that $\sigma_i = \sqrt{\lambda_i} = ||A\overrightarrow{v}_i||$ where $\overrightarrow{v}_i$ is a corresponding eigenvector.

Theorem

Given an $m \times n$ matrix A, $A^T A$ is symmetric and the eigenvectors of $A^T A$, $\{\overrightarrow{v}_1, \overrightarrow{v}_2, \ldots, \overrightarrow{v}_n\}$ form an orthonormal basis for $\mathbb{R}^n$. Assuming the eigenvalues are decreasing order and there are r non-zero values, then $\{A\overrightarrow{v}_1, A\overrightarrow{v}_2, \ldots, A\overrightarrow{v}_r\}$ is an orthogonal basis for the column space (image) of A, a subspace of $\mathbb{R}^M$.

Singular Value Decomposition: Theorem

Theorem

Let A be a $m \times n$ matrix of rank r. Then there exists an $m \times n$ matrix Σ for which the diagonal entries consist of the singular values of A, $\sigma_i = \sqrt{\lambda_i}$, an $m \times m$ orthogonal matrix U, and an $n \times n$ orthogonal matrix V such that

$$A = U \Sigma V^T.$$

Singular Value Decomposition: $A = U\Sigma V^T$

$V = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix}$ Orthogonal Matrix of Eigenvectors

$D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix}$ Singular Values $\sigma_i = \sqrt{\lambda_i} = ||A\vec{v}_i||$

$\Sigma = \begin{bmatrix} D & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix}$ $m \times n$ Diagonal Matrix

$U = \begin{bmatrix} \frac{A\vec{v}_1}{||A\vec{v}_1||} & \cdots & \frac{A\vec{v}_r}{||A\vec{v}_r||} & 0 \cdots 0 \end{bmatrix}$ $m \times m$ Matrix of Normed Images

$$\frac{A\vec{v}_i}{||A\vec{v}_i||}\sigma_i = A\vec{v}_i \implies U\Sigma = AV \text{ and } U\Sigma V^T = AVV^T = A$$

Singular Value Decomposition: $m \times n$ with $n > m$

Using our original example

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

we can write $A = U\Sigma V^T = UDV^T$, the Reduced SVD with

$$U = \begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}$$

$$\Sigma = D = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix}$$

$$V^T = \begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \end{bmatrix}$$

Singular Value Decomposition: $m \times n$ with $n < m$

If we start with a 3×2 matrix like so:

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \text{ so that } A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}$$

and we get eigenvalues of $2 \pm \sqrt{2} \approx 3.414$ and 0.586 and the matrix of eigenvectors is

$$V \approx \begin{bmatrix} 0.383 & 0.924 \\ 0.924 & -0.383 \end{bmatrix}.$$

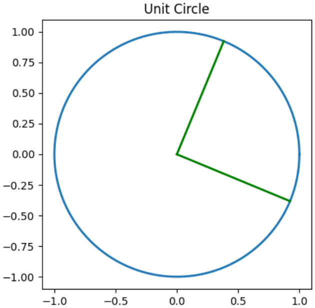
Using these we get

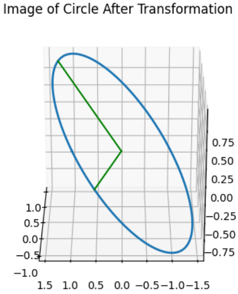
$$U = \begin{bmatrix} 1/2 & -1/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \\ 1/2 & -1/2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sqrt{2+\sqrt{2}} & 0 \\ 0 & \sqrt{2-\sqrt{2}} \end{bmatrix} \approx \begin{bmatrix} 1.84 & 0 \\ 0 & 0.765 \end{bmatrix}$$

Singular Value Decomposition: $m \times n$ with $n < m$

The result of transforming the unit circle with

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \approx \begin{bmatrix} 1/2 & -1/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \\ 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} \sqrt{2+\sqrt{2}} & 0 \\ 0 & \sqrt{2-\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0.383 & 0.924 \\ 0.924 & -0.383 \end{bmatrix}$$





Singular Value Decomposition: $m \times n$ with $n = m$

If we start with a 3×2 matrix like so:

$$A = \begin{bmatrix} 2 & 3 \\ 0 & 4 \end{bmatrix} \text{ so that } A^T A = \begin{bmatrix} 4 & 6 \\ 6 & 25 \end{bmatrix}$$

then we get eigenvalues of $\lambda_i \approx 2.41$ and 26.59 and the matrix of eigenvectors is

$$V \approx \begin{bmatrix} -0.966 & -0.257 \\ 0.257 & -0.966 \end{bmatrix}.$$

Using these we get

$$U = \begin{bmatrix} -0.750 & -0.662 \\ 0.662 & -0.750 \end{bmatrix} \qquad \Sigma \approx \begin{bmatrix} 1.551 & 0 \\ 0 & 5.157 \end{bmatrix}$$

Singular Value Decomposition: $m \times n$ with $n = m$

The result of transforming the unit circle with

$$A = \begin{bmatrix} 2 & 3 \\ 0 & 4 \end{bmatrix} \approx \begin{bmatrix} -0.750 & -0.662 \\ 0.662 & -0.750 \end{bmatrix} \begin{bmatrix} 1.551 & 0 \\ 0 & 5.157 \end{bmatrix} \begin{bmatrix} -0.966 & 0.257 \\ -0.257 & -0.966 \end{bmatrix}$$

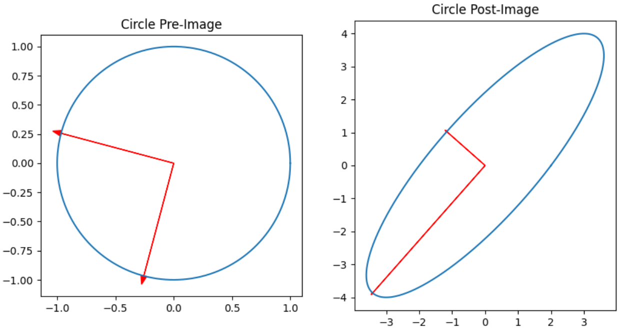


Table of Contents

1 Objectives

2 Singular Value Decomposition

3 Means and Variance

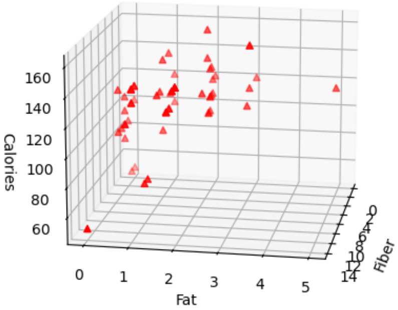
4 Principal Component Analysis

Displaying Some Data

Here we have a matrix of nutritional data for 77 different cereals:

$$Data = \begin{bmatrix} fiber \\ fat \\ calories \end{bmatrix} = \begin{bmatrix} 10.0 & 2.0 & 9.0 & \cdots & 3.0 & 3.0 & 1.0 \\ 1.0 & 5.0 & 1.0 & \cdots & 1.0 & 1.0 & 1.0 \\ 70.0 & 120.0 & 70.0 & \cdots & 100.0 & 100.0 & 110.0 \end{bmatrix}$$

And, here is a scatter plot of that data:



Constructing the Sample Mean Vector

We will begin by finding the mean for each row of data:

$$M = \begin{bmatrix} mean\ fiber \\ mean\ fat \\ mean\ calories \end{bmatrix} = \begin{bmatrix} \overline{fib} \\ \overline{fat} \\ \overline{cal} \end{bmatrix}$$

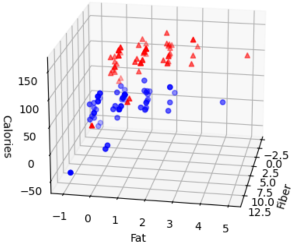
$$= \frac{1}{N} \sum_{cols\ of\ Data} Data_i = \begin{bmatrix} 2.15194805194805 \\ 1.01298701298701 \\ 106.883116883117 \end{bmatrix}$$

Constructing the Mean-Deviation Form

Then we subtract the mean values from each entry to get the signed deviations, $(x - \overline{x})$:

$$B = \begin{bmatrix} fiber - \overline{fib} \\ fat - \overline{fat} \\ calories - \overline{cal} \end{bmatrix} = \begin{bmatrix} 7.85 & -0.152 & \cdots & 0.848 & -1.15 \\ -0.013 & 3.99 & \cdots & -0.013 & -0.013 \\ -36.9 & 13.1 & \cdots & -6.88 & 3.12 \end{bmatrix}$$

If we plot this new data set, it has the same form but is centered on the mean vector:



Construct the Covariance Matrix

Next we construct the covariance matrix

$$S = \frac{1}{N-1} BB^T = \begin{bmatrix} 5.68 & 0.0401 & -13.6 \\ 0.0401 & 1.01 & 9.78 \\ -13.6 & 9.78 & 380.0 \end{bmatrix}$$

in which each entry looks like

$$\sum_i \frac{(x_i - \overline{x})^2}{N-1} \text{ or } \sum_i \frac{(x_i - \overline{x})(y_i - \overline{y})}{N-1}$$

the variance in individual variables or covariance between variables.

Variance, Covariance, and Total Variance

Definition

Given sets of data $X = \{x_1, x_2, \ldots, x_N\}$ and $Y = \{y_1, y_2, \ldots, y_N\}$ with means $\overline{x}$ and $\overline{y}$ the sums

$$\sum_i \frac{(x_i - \overline{x})^2}{N-1} \text{ and } \sum_i \frac{(y_i - \overline{y})^2}{N-1}$$

measure the variance of each individual set data from its mean. While the sum

$$\sum_i \frac{(x_i - \overline{x})(y_i - \overline{y})}{N-1}$$

measures the covariance between the variables (how they change compared to one another), if this is 0 we say they are uncorrelated. Finally the sum of the variances across variables is called the total variance, a measure of how spread out all the data is.

Variance, Covariance, and Total Variance

Returning to the matrix

$$S = \frac{1}{N-1} BB^T = \begin{bmatrix} 5.68 & 0.0401 & -13.6 \\ 0.0401 & 1.01 & 9.78 \\ -13.6 & 9.78 & 380.0 \end{bmatrix}$$

the diagonal entries are the variances, the off diagonal entries are the covariances, and the sum of the diagonal entries, called the trace, is the total variance, $trace(S) \approx 386.32$

Table of Contents

1 Objectives

2 Singular Value Decomposition

3 Means and Variance

4 Principal Component Analysis

Change of Data Variables: What We Want

Theorem (Principal Axes Theorem)

Let A be an $n \times n$ symmetric matrix, then there is an orthogonal change of variable, $\overrightarrow{x} = P\overrightarrow{y}$, that transforms the quadratic form $\overrightarrow{x}^T A\overrightarrow{x}$ into a quadratic form $\overrightarrow{y}^T D\overrightarrow{y}$ with no cross terms. Note the orthogonal matrix P and diagonal matrix D, with $A = PDP^T$, exist because A is symmetric.

Change of Data Variables: Relation to Covariance

From before we have our covariance matrix

$$S = \frac{1}{N-1} BB^T \approx \begin{bmatrix} 5.68 & 0.0401 & -13.6 \\ 0.0401 & 1.01 & 9.78 \\ -13.6 & 9.78 & 380.0 \end{bmatrix}$$

which by construction is symmetric so that we can find a change of variable with

$$\vec{x}^T S \vec{x} = \vec{y}^T D \vec{y}$$

where D is a diagonal matrix.

Constructing Principal Components: Example

As before, we find D and P so that $S = PDP^{-1} = PDP^T$ using eigenvalues and eigenvectors:

$$D \approx \begin{bmatrix} 380.0 & 0 & 0 \\ 0 & 5.22 & 0 \\ 0 & 0 & 0.726 \end{bmatrix} \text{ and } P \approx \begin{bmatrix} 0.0363 & 0.995 & -0.0874 \\ -0.0257 & 0.0883 & 0.996 \\ -0.999 & 0.0339 & -0.0288 \end{bmatrix}$$

Then we let $\vec{y} = P^T \vec{x}$ to get

$$\vec{y}^{\,T} D \vec{y} = \left(P^T \vec{x}\right)^T DP^T \vec{x} = \vec{x}^{\,T} PDP^T \vec{x} = \vec{x}^{\,T} S \vec{x}.$$

The advantage of working with the $\vec{y}$ is that pairs of variables are uncorrelated, and of course calculations with a diagonal matrix are simpler.

Constructing Principal Components: Example

Since the matrices P and P^T are orthogonal they don't change lengths or angles so

$$Total\ \ Variance = trace(S) = trace(D) \approx 386.32.$$

We can more easily look at variation due to the variables $\vec{y}$ since they are uncorrelated, in particular

- $\vec{y}_1 = \vec{P}_1^T \vec{x}$ accounts for $380/386 \approx 98.45\%$ of variation,
- $\vec{y}_2 = \vec{P}_2^T \vec{x}$ accounts for $5/386 \approx 1.30\%$ of variation, and
- $\vec{y}_3 = \vec{P}_3^T \vec{x}$ accounts for $0.73/386 \approx 0.19\%$ of variation.

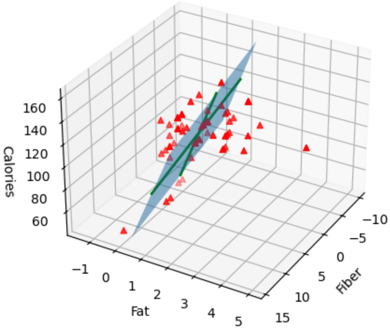
This means if we describe our data using two variables $\vec{y}_1$ and $\vec{y}_2$ instead of three separate variables, we still get approximately 99.75% of the same information.

Constructing Principal Components: Example

With

$$\overrightarrow{v} = \langle 0.036, -0.026, -0.999 \rangle \text{ and } \overrightarrow{w} = \langle 0.996, 0.088, 0.034 \rangle$$

(the eigenvectors corresponding to the two largest eigenvalues) the plane defined by $M + \overrightarrow{v}\,t + \overrightarrow{w}s$ approximates our data in two dimensions:



Constructing Principal Components: A Little Theory

The covariance matrix for the $\overrightarrow{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

Constructing Principal Components: A Little Theory

The covariance matrix for the $\overrightarrow{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

- X, Data Matrix

Constructing Principal Components: A Little Theory

The covariance matrix for the $\overrightarrow{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

- X, Data Matrix
- $\overline{X}$, Sample Means

Constructing Principal Components: A Little Theory

The covariance matrix for the $\overrightarrow{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

- X, Data Matrix
- $\overline{X}$, Sample Means
- $B = X - \overline{X}$, Mean Deviation

Constructing Principal Components: A Little Theory

The covariance matrix for the $\vec{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

- X, Data Matrix
- $\overline{X}$, Sample Means
- $B = X - \overline{X}$, Mean Deviation
- $S = \frac{1}{N-1}BB^T$, Covariance

Constructing Principal Components: A Little Theory

The covariance matrix for the $\overrightarrow{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

- X, Data Matrix
- $\overline{X}$, Sample Means
- $B = X - \overline{X}$, Mean Deviation
- $S = \frac{1}{N-1} BB^T$, Covariance
- $S = PDP^T$, Ortho. Diag.

Constructing Principal Components: A Little Theory

The covariance matrix for the $\overrightarrow{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

- X, Data Matrix
- $\overline{X}$, Sample Means
- $B = X - \overline{X}$, Mean Deviation
- $S = \frac{1}{N-1}BB^T$, Covariance
- $S = PDP^T$, Ortho. Diag.
- Define $Y = P^T X$,

Constructing Principal Components: A Little Theory

The covariance matrix for the $\overrightarrow{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

- X, Data Matrix
- $\overline{X}$, Sample Means
- $B = X - \overline{X}$, Mean Deviation
- $S = \frac{1}{N-1}BB^T$, Covariance
- $S = PDP^T$, Ortho. Diag.
- Define $Y = P^T X$,

$$\begin{aligned} \overline{Y} &= \frac{1}{N}\sum_i Y_i \\ &= \frac{1}{N}\sum P^T X_i \\ &= P^T\left(\frac{1}{N}\sum X_i\right) \\ &= P^T\overline{X} \end{aligned}$$

Constructing Principal Components: A Little Theory

The covariance matrix for the $\overrightarrow{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

- X, Data Matrix
- $\overline{X}$, Sample Means
- $B = X - \overline{X}$, Mean Deviation
- $S = \frac{1}{N-1}BB^T$, Covariance
- $S = PDP^T$, Ortho. Diag.
- Define $Y = P^T X$,
- $\overline{Y} = P^T \overline{X}$

Constructing Principal Components: A Little Theory

The covariance matrix for the $\overrightarrow{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

- X, Data Matrix
- $\overline{X}$, Sample Means
- $B = X - \overline{X}$, Mean Deviation
- $S = \frac{1}{N-1} BB^T$, Covariance
- $S = PDP^T$, Ortho. Diag.
- Define $Y = P^T X$,
- $\overline{Y} = P^T \overline{X}$

$$\begin{aligned} Y - \overline{Y} &= P^T X - P^T \overline{X} \\ &= P^T \left(X - \overline{X}\right) \\ &= P^T B \end{aligned}$$

Constructing Principal Components: A Little Theory

The covariance matrix for the $\overrightarrow{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

- X, Data Matrix
- $\overline{X}$, Sample Means
- $B = X - \overline{X}$, Mean Deviation
- $S = \frac{1}{N-1}BB^T$, Covariance
- $S = PDP^T$, Ortho. Diag.
- Define $Y = P^T X$,
- $\overline{Y} = P^T\overline{X}$
- $Y - \overline{Y} = P^T B$

Constructing Principal Components: A Little Theory

The covariance matrix for the $\overrightarrow{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

- X, Data Matrix
- $\overline{X}$, Sample Means
- $B = X - \overline{X}$, Mean Deviation
- $S = \frac{1}{N-1}BB^T$, Covariance
- $S = PDP^T$, Ortho. Diag.
- Define $Y = P^T X$,
- $\overline{Y} = P^T \overline{X}$
- $Y - \overline{Y} = P^T B$

$$\begin{aligned}
\frac{1}{N-1}\left(Y - \overline{Y}\right)\left(Y - \overline{Y}\right)^T &= \frac{1}{N-1}\left(P^T B\right)\left(P^T B\right)^T \\
&= \frac{1}{N-1}P^T BB^T P \\
&= P^T \frac{1}{N-1}BB^T P \\
&= P^T SP = D
\end{aligned}$$

Constructing Principal Components: A Little Theory

The covariance matrix for the $\overrightarrow{y}$'s as defined above is precisely the diagonal matrix D of eigenvalues for the covariance matrix $S = \frac{1}{N-1}(X - \overline{X})(X - \overline{X})^T$.

- X, Data Matrix
- $\overline{X}$, Sample Means
- $B = X - \overline{X}$, Mean Deviation
- $S = \frac{1}{N-1}BB^T$, Covariance
- $S = PDP^T$, Ortho. Diag.
- Define $Y = P^T X$,
- $\overline{Y} = P^T \overline{X}$
- $Y - \overline{Y} = P^T B$
- Covariance of Y is D

$$\begin{aligned} \frac{1}{N-1}\left(Y - \overline{Y}\right)\left(Y - \overline{Y}\right)^T &= \frac{1}{N-1}\left(P^T B\right)\left(P^T B\right)^T \\ &= \frac{1}{N-1}P^T BB^T P \\ &= P^T \frac{1}{N-1}BB^T P \\ &= P^T SP = D \end{aligned}$$

Lectures on Multivariable Mathematics: Principal Component Analysis and Singular Value Decomposition

Dr. Chuck Rocca
roccac@wcsu.edu