The Math of PCA

DATASCI 347 MACHINE LEARNING 1

ALEXANDER WILLIAMS TOLBERT

Why Do We Need PCA?

The problem:

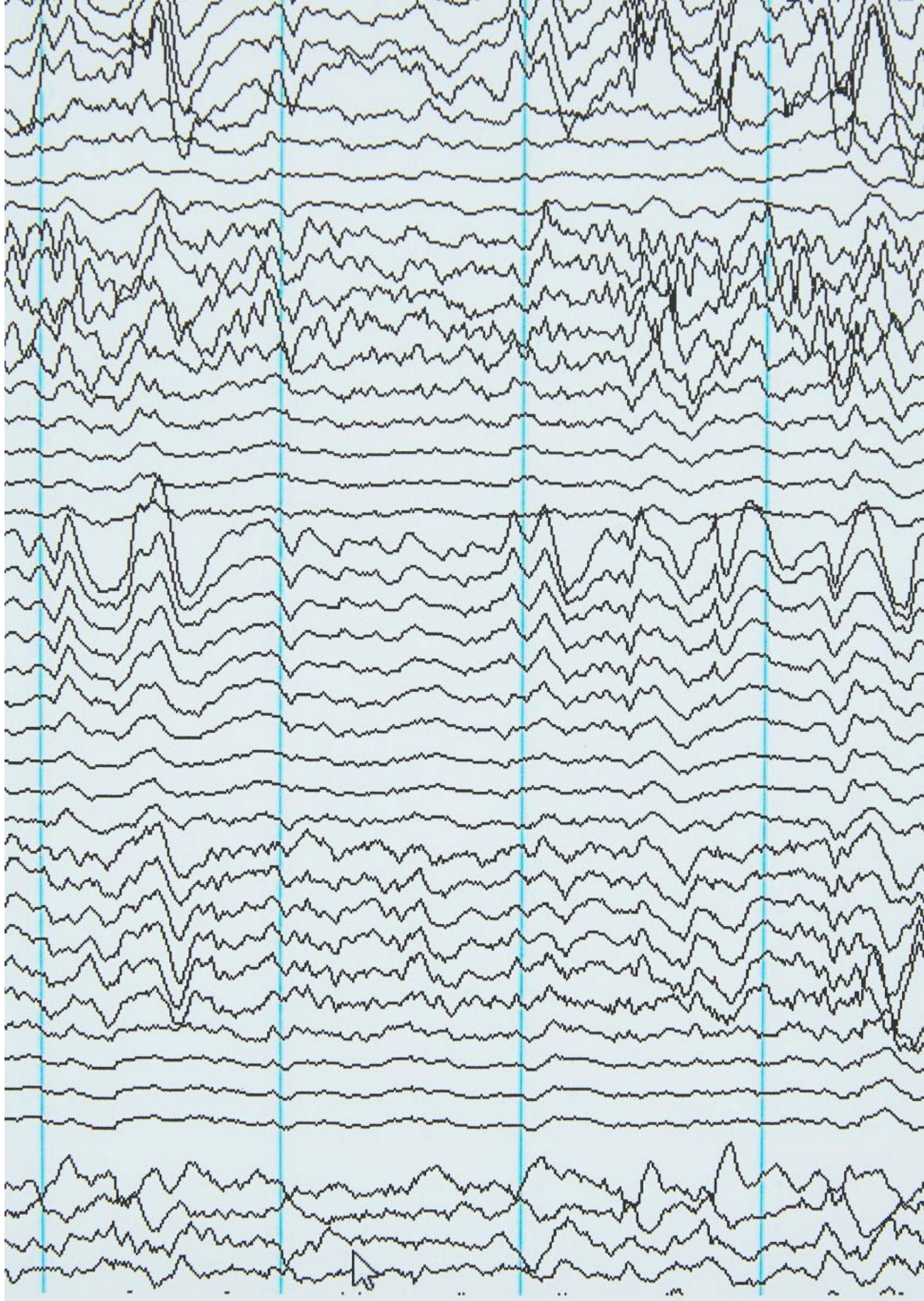
- Many datasets have **lots of variables** (dimensions).
- These variables are often **correlated** (carry overlapping info).
- High-dimensional data is hard to **analyze, visualize, and interpret**.

The solution:

- Find a smaller set of variables that still captures most of the important variation in the data.

Example:

- Systolic and diastolic blood pressure are strongly correlated.
- Do we really need both, or can we capture the information with fewer variables?



Why Does PCA Work?

PCA = **Principal Component Analysis**.

PCA creates **new variables** (principal components) that are:

- **Linear combinations** of the original variables.
- **Uncorrelated** with each other.
- **Ordered by variance explained**.

The first principal component (PC1):

- Points in the direction of **maximum variance** in the data.

The second (PC2):

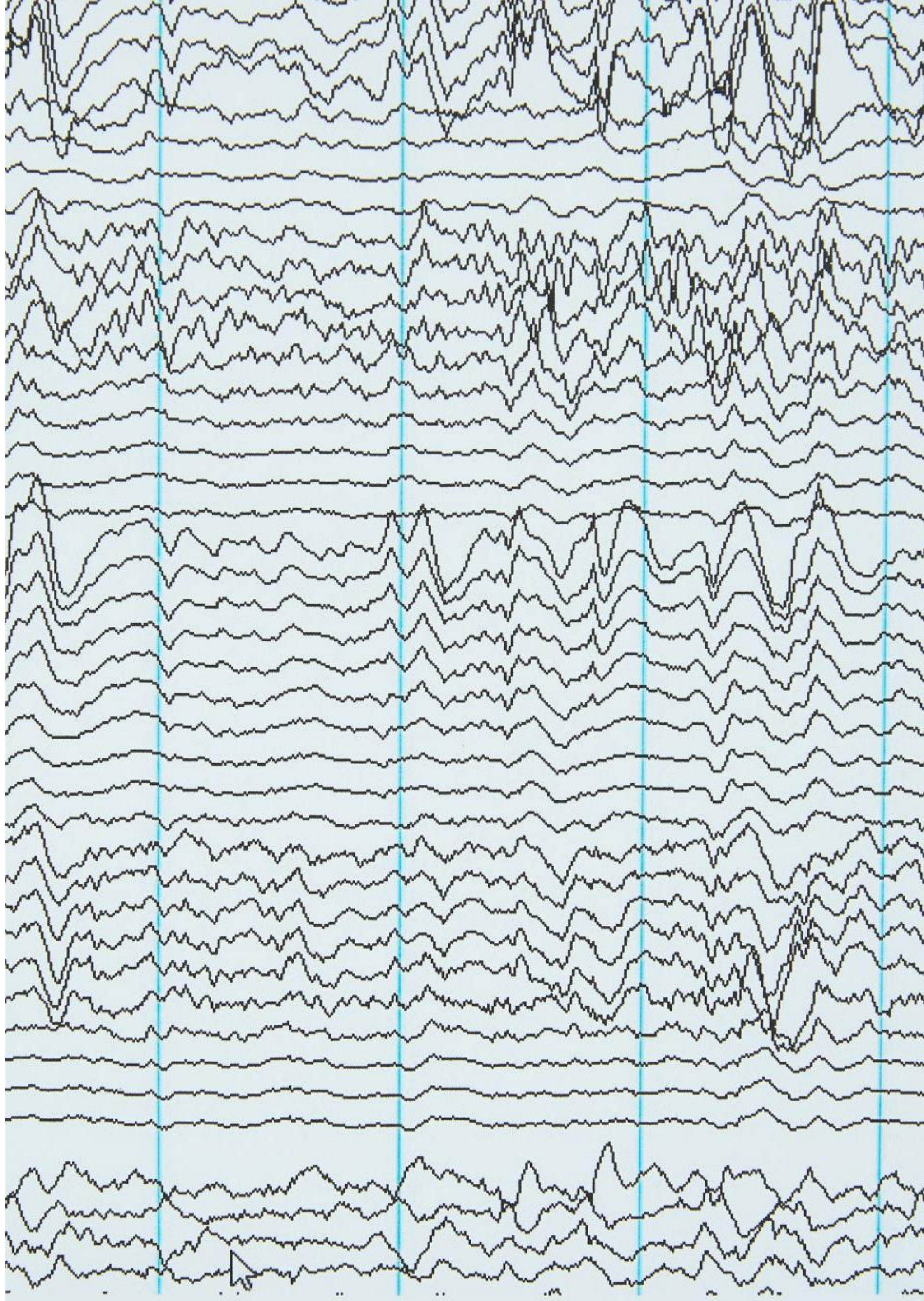
- Orthogonal to PC1.
- Explains the next largest variance.
- And so on.

Intuition Behind PCA

- Imagine your data as a **cloud of points**.
- Not all directions are equally “spread out.”
- PCA rotates the coordinate system to line up with the directions of greatest spread.

Why variance?

- Variance measures how much the data differs along a direction.
- A direction with **high variance** shows strong patterns in the data.
- A direction with **low variance** usually reflects noise or redundancy.



Step 1 – Representing the Data

- Start with a dataset as a **matrix**:
  - Rows = observations (e.g., people).
  - Columns = variables (e.g., systolic & diastolic BP).
- Example structure:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix}$$

- First step in PCA: **center the data** (subtract the mean of each column).

Step 1 – Representing the Data

- Start with a dataset as a **matrix**:
  - Rows = observations (e.g., people).
  - Columns = variables (e.g., systolic & diastolic BP).
- Example structure:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix}$$

- where n = number of observations, ddd = number of variables.
- First step in PCA: **center the data** (subtract the mean of each column).

Step 2 – Covariance Matrix

- **Covariance measures how variables change together.**
- Positive covariance → variables increase together.
- Negative covariance → one increases while the other decreases.
- Zero covariance → no linear relationship.
- Covariance matrix for d variables:

$$\Sigma = \frac{1}{n-1} X^\top X$$

- Properties:
  - Symmetric ($\Sigma = \Sigma^\top$).
  - Diagonal entries = variances of each variable.
  - Off-diagonal entries = covariances between variables.

Step 3- Variance Maximization Problem

- Goal: Find a direction v (a vector) that **maximizes variance of projected data.**
- Project data onto v:

$$z = Xv$$

- Variance of projection:

$$\mathrm{Var}(z) = v^\top \Sigma v$$

- Optimization problem:

$$\text{maximize } v^\top \Sigma v \quad \text{subject to } \|v\| = 1$$

Variance Maximization Problem

- Goal: Find a direction v (a vector) that **maximizes variance of projected data.**
- Project data onto v:

$$z = Xv$$

- Variance of projection:

$$\mathrm{Var}(z) = v^\top \Sigma v$$

- Optimization problem:

$$\text{maximize } v^\top \Sigma v \quad \text{subject to } \|v\| = 1$$

Step 4 – The Eigenvalue Problem

- The optimization problem leads to:

$$\Sigma v = \lambda v$$

- Where:
  - v = **eigenvector** (principal component direction).
  - λ = **eigenvalue** (variance explained along that direction).
- Key facts:
  - Eigenvectors of Σ give the principal component directions.
  - Eigenvalues tell us how much variance each component explains.
  - Since Σ is symmetric, eigenvectors are guaranteed to be orthogonal.

PCA via SVD (Optional / Computational Approach)

Another way to find principal components is **Singular Value Decomposition (SVD)**. For centered data X:

$$X = U\Sigma V^{\top}$$

- Columns of V are the principal components.
- Squared singular values σ_i^2 give the variance explained by each component.
- Project data onto the top k PCs:

$$X_{\text{reduced}} = XV_k$$

SVD rotates the data to align with directions of maximum spread, giving the same result as eigenvectors, often more stably.

Step 5 – Ordering and Reducing Dimensions

- Sort eigenvalues:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$$

- Order corresponding eigenvectors the same way.
- **First k components** (those with largest eigenvalues) capture the most variance.
- Dimensionality reduction:
  - Keep only the top k eigenvectors.
  - Project data:

$$X_{\text{reduced}} = XV_k$$

where V_k contains the top k eigenvectors.

Eigen-decomposition of covariance matrix

Goal: Find directions where data spreads most (principal components).

1. **Start with data** X (rows = observations, columns = variables) and center it by subtracting the mean.
2. Compute **covariance matrix**

$$\Sigma = E[(X - \bar{X})(X - \bar{X})^\top]$$

where $E[\cdot]$ = average over observations; measures how variables vary together.

3. **Eigen-decomposition:**

$$\Sigma = V \Lambda V^\top$$

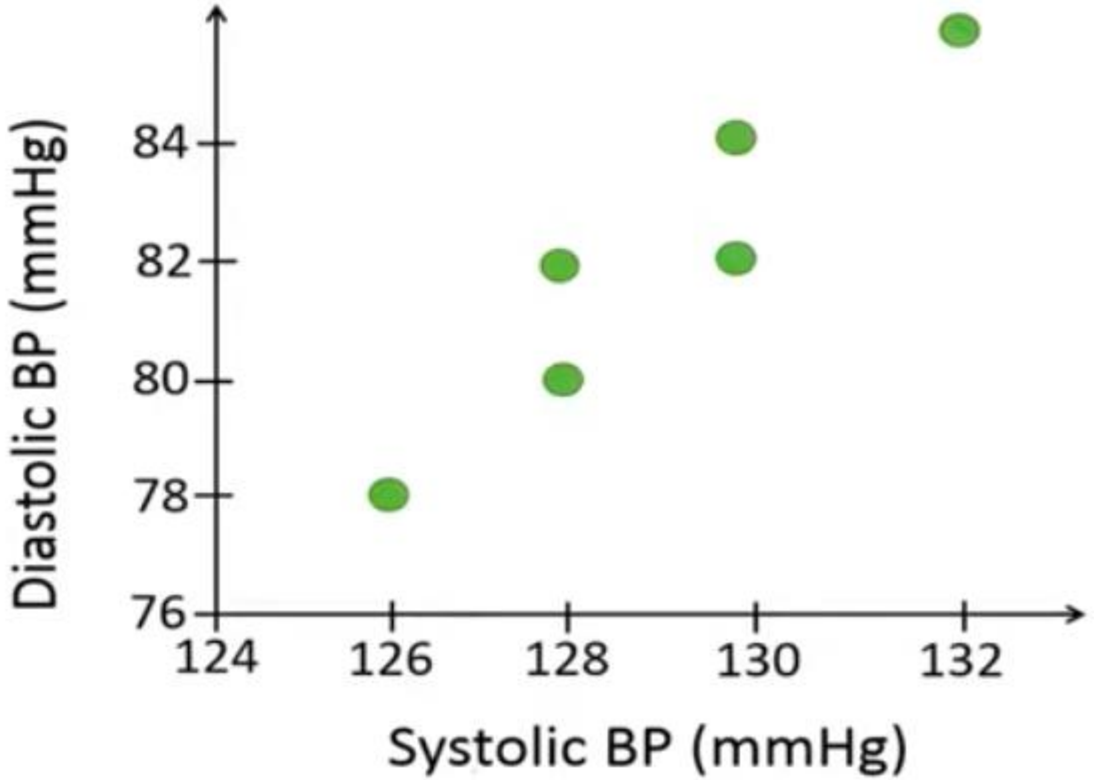
- Columns of V = orthogonal directions of maximal spread
- Diagonal Λ = variance along each direction
- Eigenvectors normalized ($\|v\| = 1$) → eigenvalue = exact variance along that direction
- $V^\top$ rotates covariance into PC axes; multiplying by V rotates back

4. **Dimensionality reduction:** Keep top k eigenvectors → main patterns, discard noise

Note: SVD can compute PCA directly from X, more stable numerically, but eigen-decomposition shows the math clearly.

Intuition: Imagine a cloud of points; eigenvectors = axes it stretches most; $V^\top$ aligns axes, V rotates back.

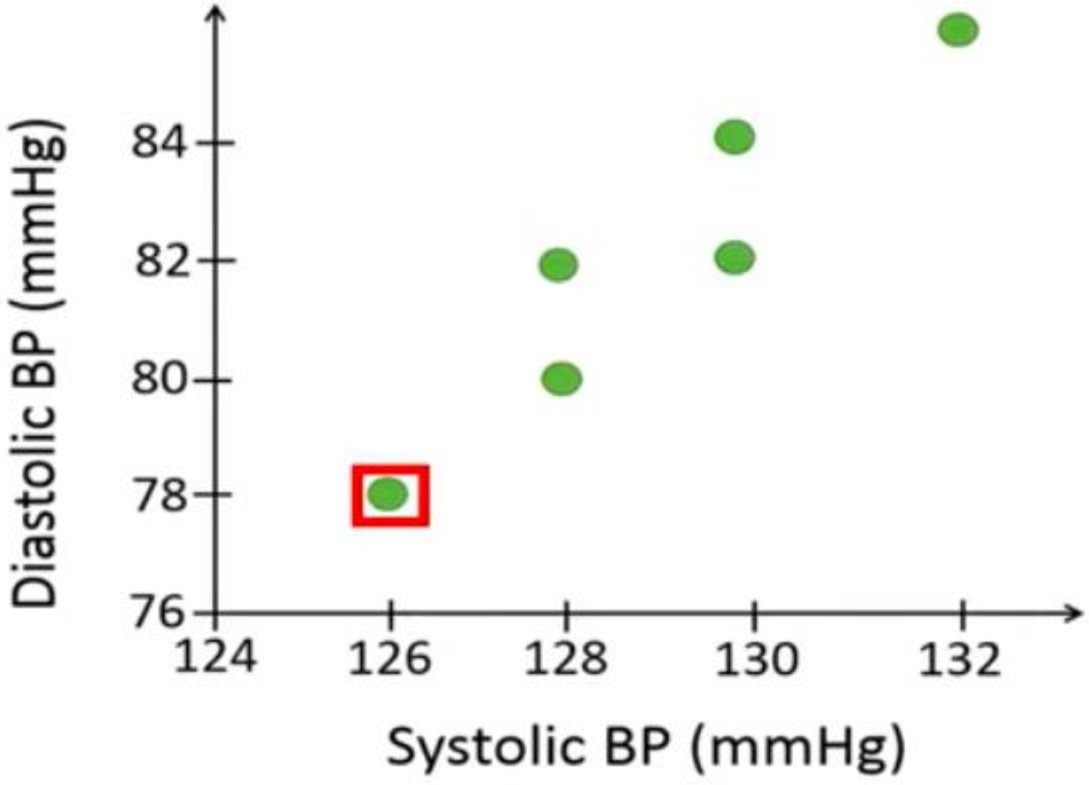
Example data



Systolic BP	Diastolic BP
126	78
128	80
128	82
130	82
130	84
132	86

To explain how the PCA works, we will use the following example data. We will use PCA to combine the two blood pressure variables into just one variable based on data from six individuals.

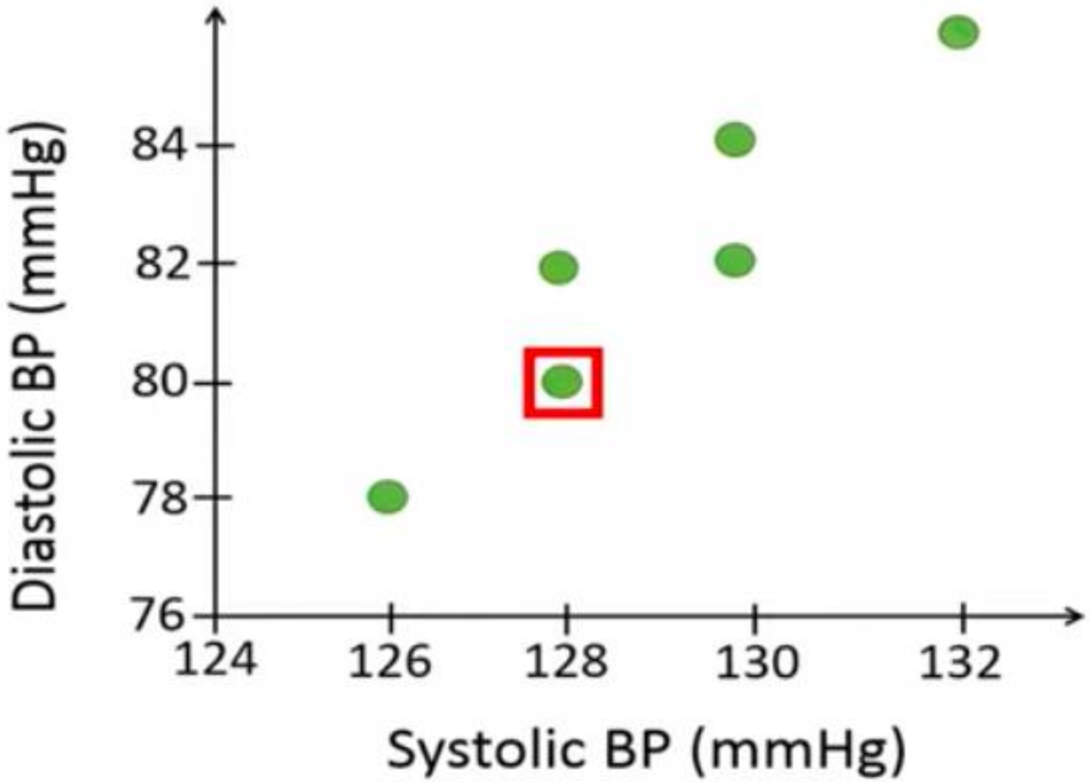
Example data



Systolic BP	Diastolic BP
126	78
128	80
128	82
130	82
130	84
132	86

For example, person number one has a diastolic blood pressure of 78 and a systolic blood pressure of 126,

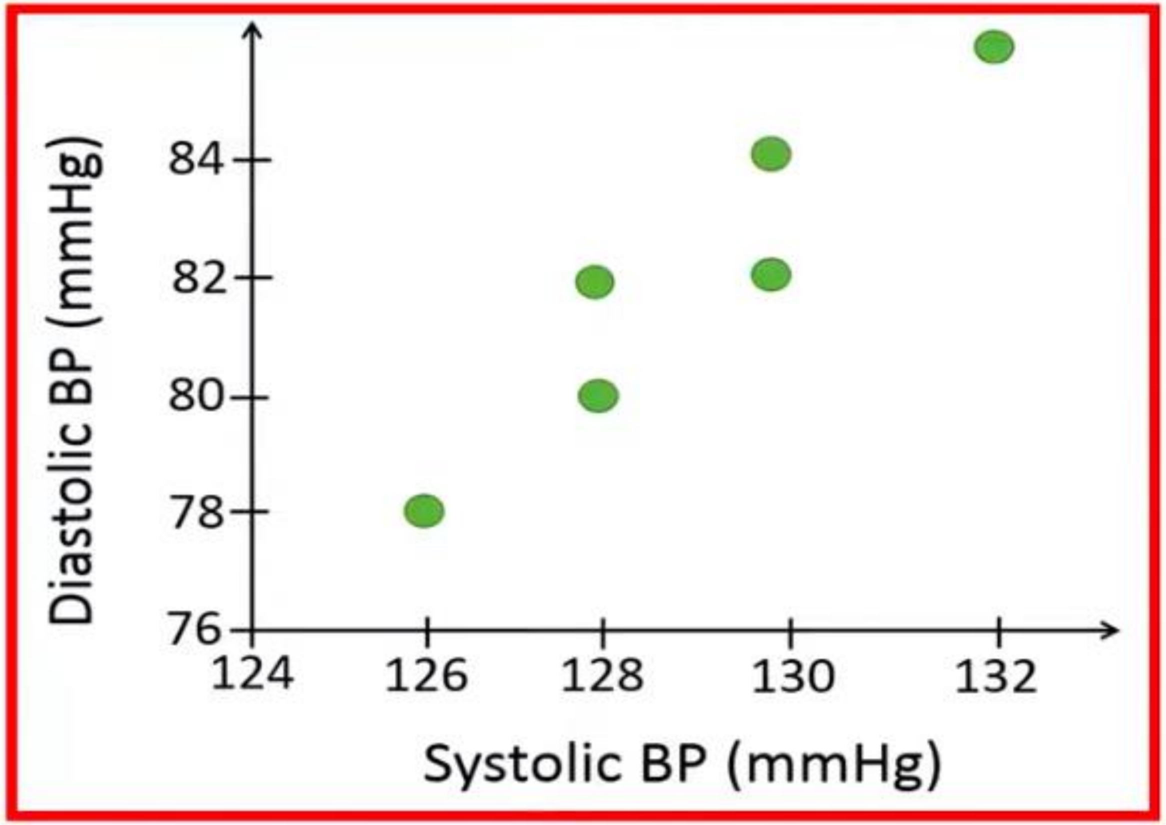
Example data



Systolic BP	Diastolic BP
126	78
128	80
128	82
130	82
130	84
132	86

whereas person number two has a diastolic blood pressure of 80 and a systolic blood pressure of 128, and so on.

Example data



Systolic BP	Diastolic BP
126	78
128	80
128	82
130	82
130	84
132	86

For this data set, it seems to be a strong positive correlation between the two variables.

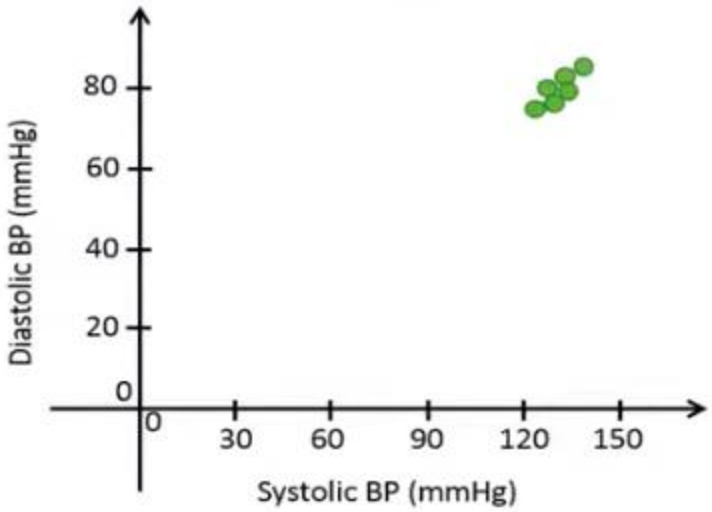
PCA

1. Center the data
2. Calculate the covariance matrix
3. Calculate eigenvalues of the covariance matrix
4. Calculate eigenvectors of the covariance matrix
5. Order the eigenvectors
6. Calculate the principal components

To compute a PCA, we can perform the following steps,

1. Center the data

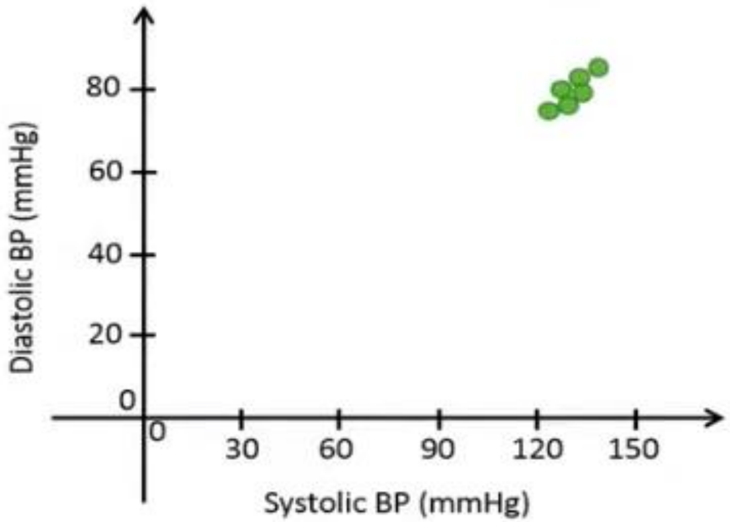
Systolic BP	Diastolic BP
126	78
128	80
128	82
130	82
130	84
132	86



Usually, one starts to center or standardize the data in the first step of the PCA analysis. In this case, we will only center the data, which means that we subtract all the values for each variable by its corresponding mean.

1. Center the data

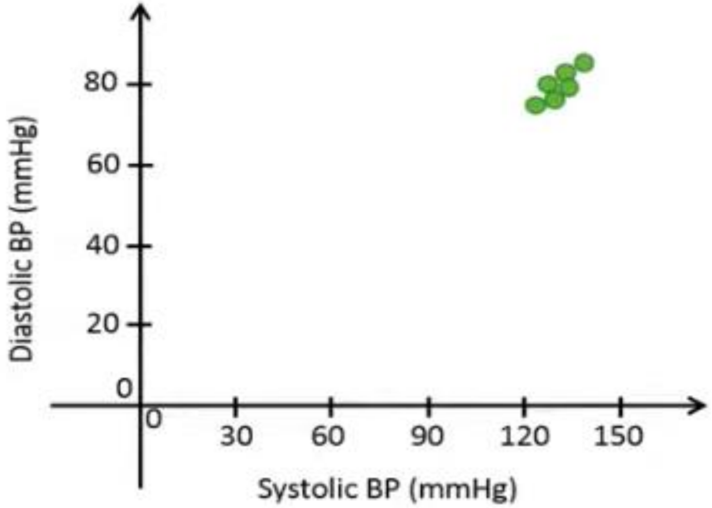
Systolic BP	Diastolic BP
126 -129 = -3	78
128 -129 = -1	80
128 -129 = -1	82
130 -129 = 1	82
130 -129 = 1	84
132 -129 = 3	86



We therefore subtract the mean systolic blood pressure,

1. Center the data

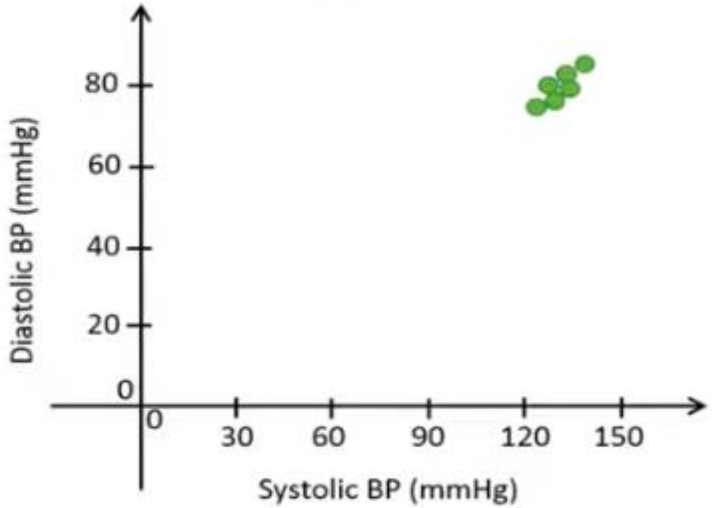
Systolic BP	Diastolic BP
126 -129 = -3	78
128 -129 = -1	80
128 -129 = -1	82
130 -129 = 1	82
130 -129 = 1	84
132 -129 = 3	86



from the individual observations.

1. Center the data

Systolic BP		Diastolic BP
126 -129 =	-3	78
128 -129 =	-1	80
128 -129 =	-1	82
130 -129 =	1	82
130 -129 =	1	84
132 -129 =	3	86

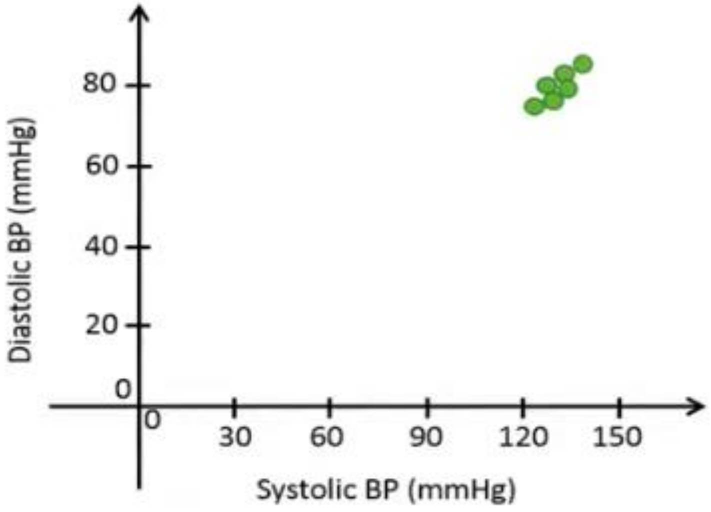


Centering the systolic blood pressure results in the following values, which tell how far away the original values are from the mean.

1. Center the data

Systolic BP	Diastolic BP
126 -129 = -3	78 -82 = -4
128 -129 = -1	80 -82 = -2
128 -129 = -1	82 -82 = 0
130 -129 = 1	82 -82 = 0
130 -129 = 1	84 -82 = 2
132 -129 = 3	86 -82 = 4

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4

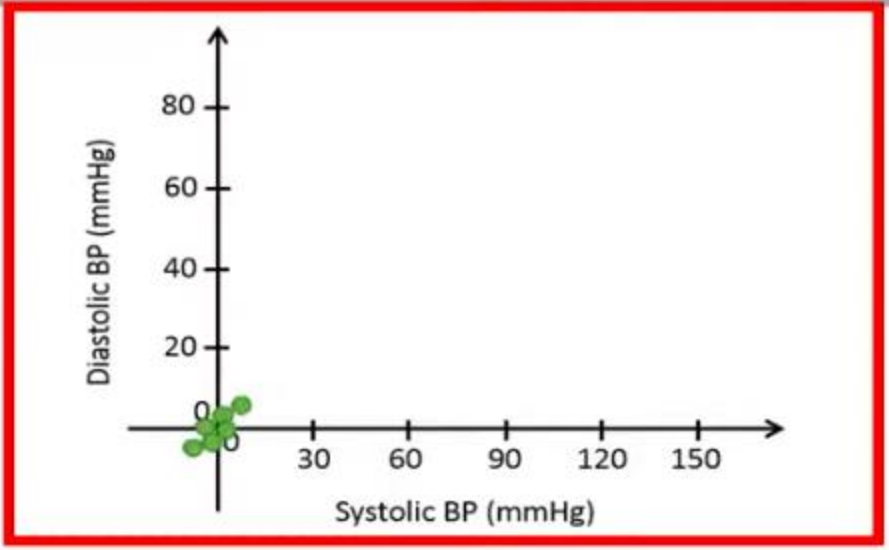


We can summarize the centered data in the following table.

1. Center the data

Systolic BP	Diastolic BP
126 -129 = -3	78 -82 = -4
128 -129 = -1	80 -82 = -2
128 -129 = -1	82 -82 = 0
130 -129 = 1	82 -82 = 0
130 -129 = 1	84 -82 = 2
132 -129 = 3	86 -82 = 4

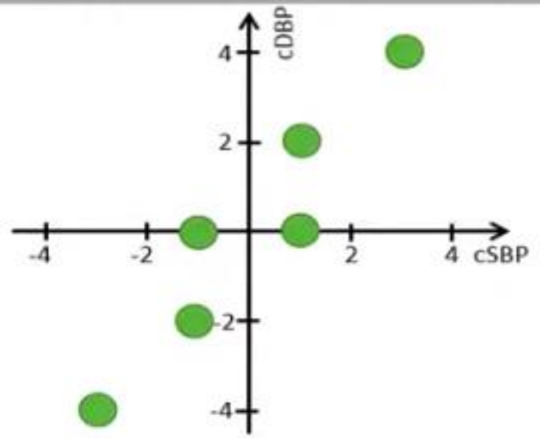
Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



When we center the data, it means that we center the data points around the origin. Centering the data around the origin will help us later when we will rotate the data.

1. Center the data

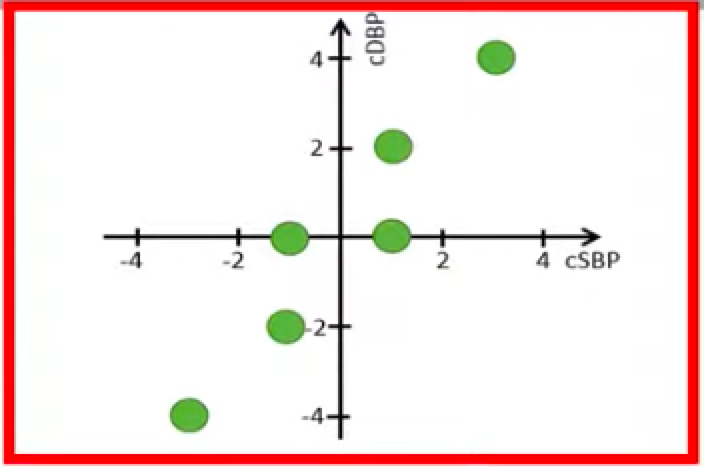
Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



After we have centered the data, we will have the following values,

1. Center the data

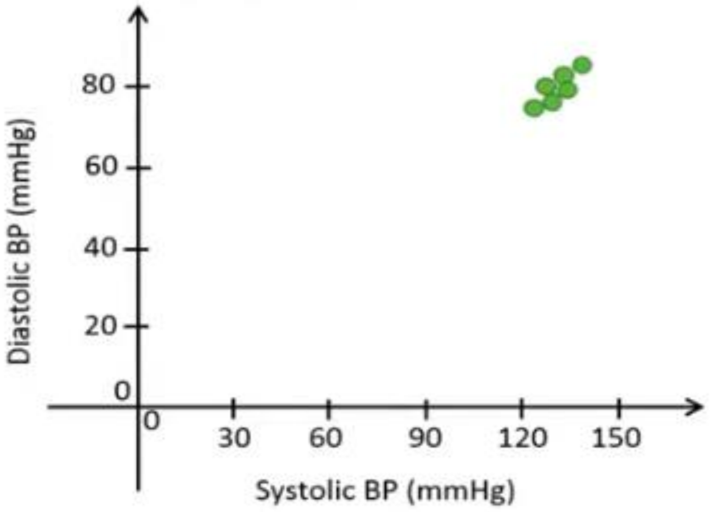
Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



which can be plotted like this, where the x-axis now represents the centered systolic blood pressure, whereas the y-axis represents the centered diastolic blood pressure.

1. Center the data

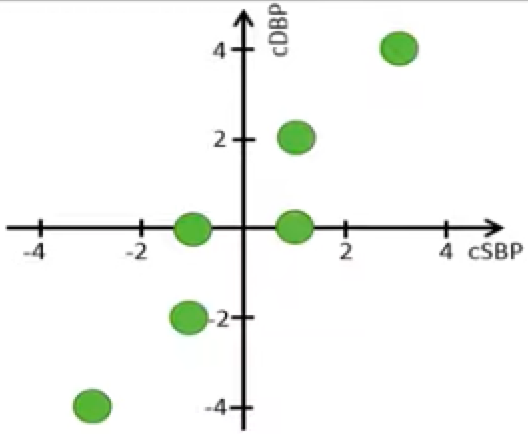
Systolic BP	Diastolic BP
126 -129 = -3	78 -82 = -4
128 -129 = -1	80 -82 = -2
128 -129 = -1	82 -82 = 0
130 -129 = 1	82 -82 = 0
130 -129 = 1	84 -82 = 2
132 -129 = 3	86 -82 = 4



We then do the same calculations for the diastolic blood pressure, which has a mean value of 82.

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4

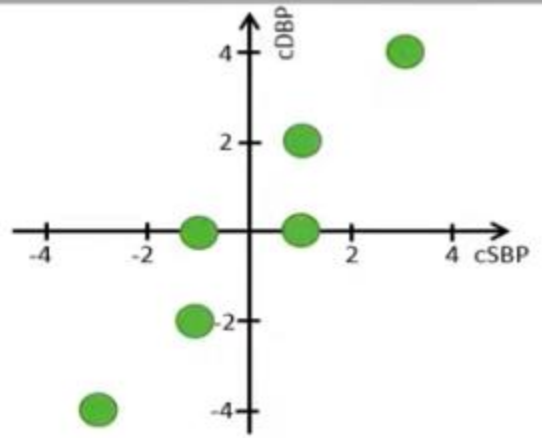


	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

Next, we calculate the covariance matrix based on the centered data. Note that we would have got the same values in the covariance matrix if we instead would have used the original data since the variance does not change when we center the data.

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4

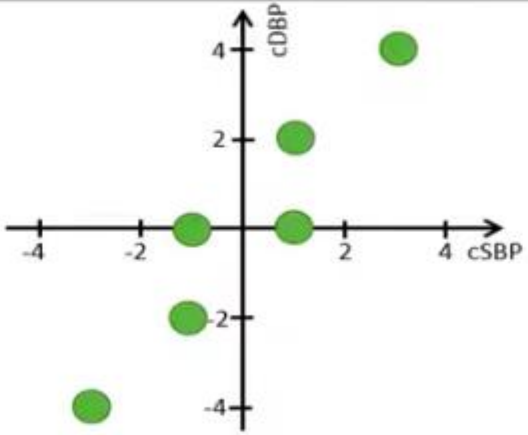


	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

Remember that the main diagonal of the covariance matrix includes the variance of each variable.

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



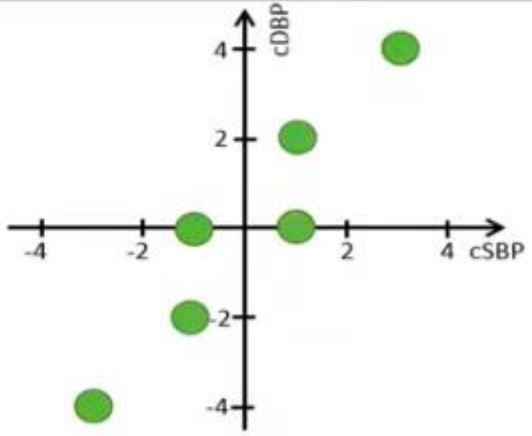
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\boxed{\mathrm{var(cSBP)} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{cSBP}_i - \overline{\mathrm{cSBP}})^2} = ((-3)^2 + (-1)^2 + (-1)^2 + 1^2 + 1^2 + 3^2)/(6-1) = 22/5 = 4.4$$

The sample variance of the centered systolic blood pressure is calculated like this.

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



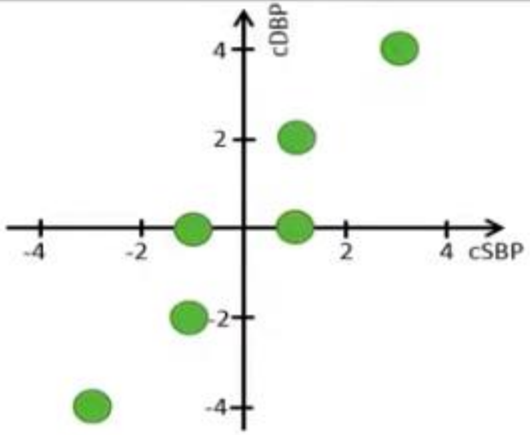
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\mathrm{var}(\mathrm{cSBP}) = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{cSBP}_i - \boxed{\overline{\mathrm{cSBP}}})^2 = ((-3)^2 + (-1)^2 + (-1)^2 + 1^2 + 1^2 + 3^2)/(6-1) = 22/5 = 4.4$$

When we calculate the variance of the centered data, the calculations become a bit simpler since the mean of the centered data is always equal to zero.

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



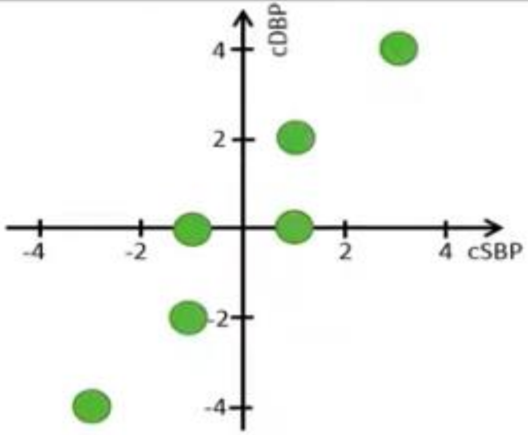
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\mathrm{var}(cSBP) = \frac{1}{n-1}\sum_{i=1}^{n}(cSBP_i - \overline{cSBP})^2 = \boxed{(-3)^2 + (-1)^2 + (-1)^2 + 1^2 + 1^2 + 3^2} / (6-1) = 22/5 = 4.4$$

To calculate the sample variance of the centered data, we therefore simply sum the squared values,

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



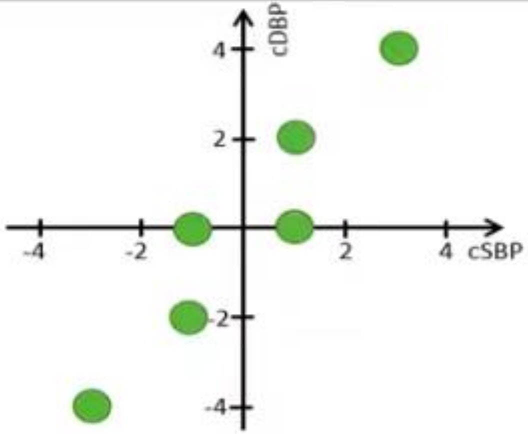
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\mathrm{var(cSBP)} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{cSBP}_i - \overline{\mathrm{cSBP}})^2 = ((-3)^2 + (-1)^2 + (-1)^2 + 1^2 + 1^2 + 3^2)\boxed{/\,(6-1)} = 22/5 = 4.4$$

and divide by the sample size minus one.

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



	SBP	DBP
SBP	4.4	5.6
DBP	5.6	$\boxed{8.0}$

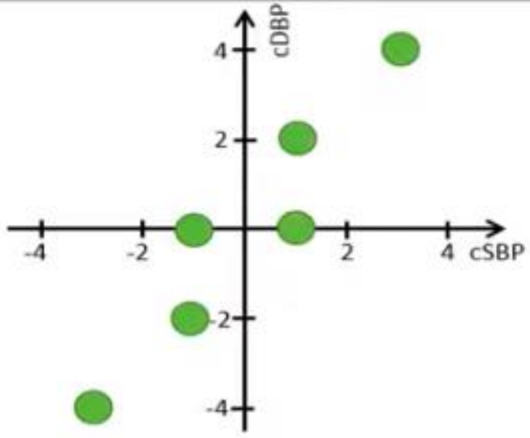
$$\mathrm{var(cSBP)} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{cSBP}_i - \overline{\mathrm{cSBP}})^2 = ((-3)^2 + (-1)^2 + (-1)^2 + 1^2 + 1^2 + 3^2)/(6-1) = 22/5 = 4.4$$

$$\boxed{\mathrm{var(cDBP)} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{cDBP}_i - \overline{\mathrm{cDBP}})^2 = ((-4)^2 + (-2)^2 + 0^2 + 0^2 + 2^2 + 4^2)/(6-1) = 40/5 = 8}$$

Then we calculate the variance of the diastolic blood pressure by using the same equation.

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



	SBP	DBP
SBP	4.4	$\boxed{5.6}$
DBP	$\boxed{5.6}$	8.0

$$\mathrm{var(cSBP)} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{cSBP}_i - \overline{\mathrm{cSBP}})^2 = ((-3)^2+(-1)^2+(-1)^2+1^2+1^2+3^2)/(6-1) = 22/5 = 4.4$$

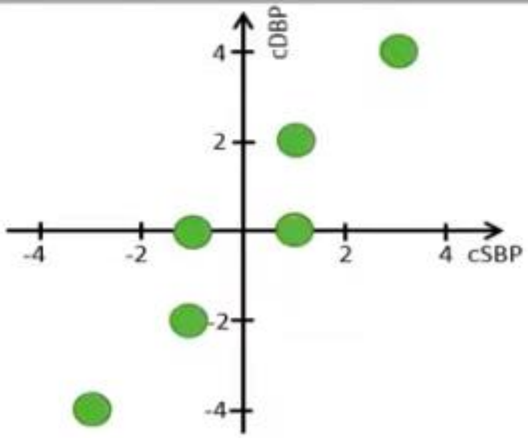
$$\mathrm{var(cDBP)} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{cDBP}_i - \overline{\mathrm{cDBP}})^2 = ((-4)^2+(-2)^2+0^2+0^2+2^2+4^2)/(6-1) = 40/5 = 8$$

$$\boxed{\mathrm{cov(cSBP,cDBP)} = \frac{1}{n-1}\sum(\mathrm{cSBP}_i - \overline{\mathrm{cSBP}})\cdot(\mathrm{cDBP}_i - \overline{\mathrm{cDBP}}) = ((-3)\cdot(-4)+(-1)\cdot(-2)+(-1)\cdot 0+1\cdot 0+1\cdot 2+3\cdot 4)/(6-1) = 28/5 = 5.6}$$

Finally, we calculate the covariance, which is a measure of how much the two variables spread together.

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\mathrm{var}(cSBP) = \frac{1}{n-1}\sum_{i=1}^{n}(cSBP_i - \overline{cSBP})^2 = ((-3)^2 + (-1)^2 + (-1)^2 + 1^2 + 1^2 + 3^2)/(6-1) = 22/5 = 4.4$$

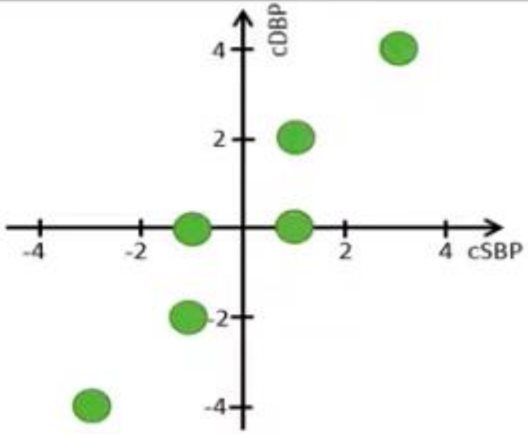
$$\mathrm{var}(cDBP) = \frac{1}{n-1}\sum_{i=1}^{n}(cDBP_i - \overline{cDBP})^2 = ((-4)^2 + (-2)^2 + 0^2 + 0^2 + 2^2 + 4^2)/(6-1) = 40/5 = 8$$

$$\mathrm{cov}(cSBP, cDBP) = \frac{1}{n-1}\sum \boxed{(cSBP_i - \overline{cSBP})\cdot(cDBP_i - \overline{cDBP})} = ((-3)\cdot(-4) + (-1)\cdot(-2) + (-1)\cdot 0 + 1\cdot 0 + 1\cdot 2 + 3\cdot 4)/(6-1) = 28/5 = 5.6$$

The sample covariance is calculated by multiplying the centered values of the two variables.

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\mathrm{var}(\mathrm{cSBP}) = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{cSBP}_i - \overline{\mathrm{cSBP}})^2 = ((-3)^2 + (-1)^2 + (-1)^2 + 1^2 + 1^2 + 3^2)/(6-1) = 22/5 = 4.4$$

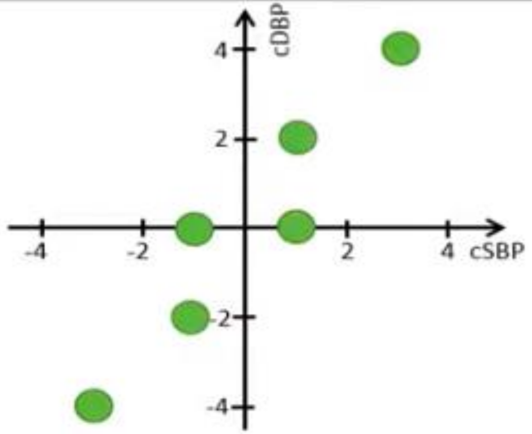
$$\mathrm{var}(\mathrm{cDBP}) = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{cDBP}_i - \overline{\mathrm{cDBP}})^2 = ((-4)^2 + (-2)^2 + 0^2 + 0^2 + 2^2 + 4^2)/(6-1) = 40/5 = 8$$

$$\mathrm{cov}(\mathrm{cSBP},\mathrm{cDBP}) = \frac{1}{n-1}\sum(\mathrm{cSBP}_i - \overline{\mathrm{cSBP}})\cdot(\mathrm{cDBP}_i - \overline{\mathrm{cDBP}}) = (\boxed{(-3)\cdot(-4)} + (-1)\cdot(-2) + (-1)\cdot 0 + 1\cdot 0 + 1\cdot 2 + 3\cdot 4)/(6-1) = 28/5 = 5.6$$

For example, we multiply the centered values for person number one,

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\mathrm{var(cSBP)} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{cSBP}_i - \overline{\mathrm{cSBP}})^2 = ((-3)^2 + (-1)^2 + (-1)^2 + 1^2 + 1^2 + 3^2)/(6-1) = 22/5 = 4.4$$

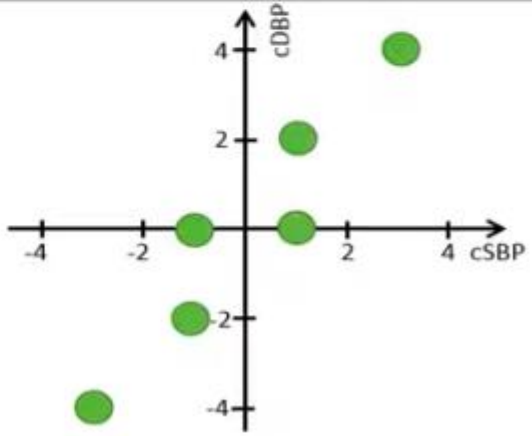
$$\mathrm{var(cDBP)} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{cDBP}_i - \overline{\mathrm{cDBP}})^2 = ((-4)^2 + (-2)^2 + 0^2 + 0^2 + 2^2 + 4^2)/(6-1) = 40/5 = 8$$

$$\mathrm{cov(cSBP,cDBP)} = \frac{1}{n-1}\sum(\mathrm{cSBP}_i - \overline{\mathrm{cSBP}})\cdot(\mathrm{cDBP}_i - \overline{\mathrm{cDBP}}) = ((-3)\cdot(-4\boxed{+(-1)\cdot(-2)+}(-1)\cdot 0 + 1\cdot 0 + 1\cdot 2 + 3\cdot 4)/(6-1) = 28/5 = 5.6$$

and add that to the product of the centered values for person number two, and so on.

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\text{var}(\text{cSBP}) = \frac{1}{n-1}\sum_{i=1}^{n}(\text{cSBP}_i - \overline{\text{cSBP}})^2 = ((-3)^2+(-1)^2+(-1)^2+1^2+1^2+3^2)/(6-1) = 22/5 = 4.4$$

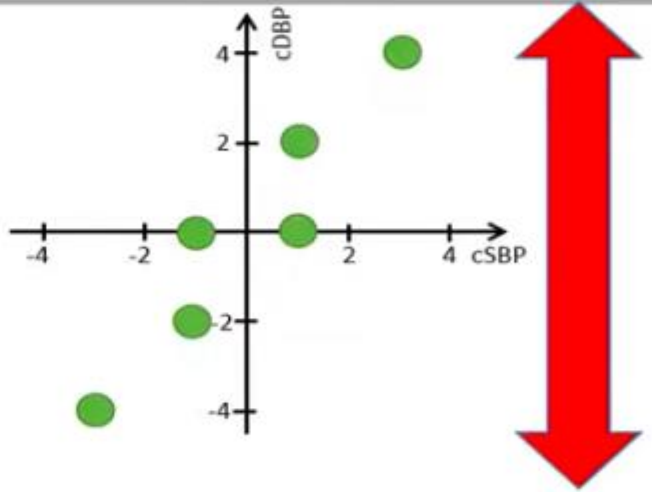
$$\text{var}(\text{cDBP}) = \frac{1}{n-1}\sum_{i=1}^{n}(\text{cDBP}_i - \overline{\text{cDBP}})^2 = ((-4)^2+(-2)^2+0^2+0^2+2^2+4^2)/(6-1) = 40/5 = 8$$

$$\text{cov}(\text{cSBP},\text{cDBP}) = \frac{1}{n-1}\sum(\text{cSBP}_i - \overline{\text{cSBP}})\cdot(\text{cDBP}_i - \overline{\text{cDBP}}) = ((-3)\cdot(-4)+(-1)\cdot(-2)+(-1)\cdot 0+1\cdot 0+1\cdot 2+3\cdot 4)/\boxed{(6-1)} = 28/5 = 5.6$$

Finally, we divide the sum of the products by the sample size minus one.

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4

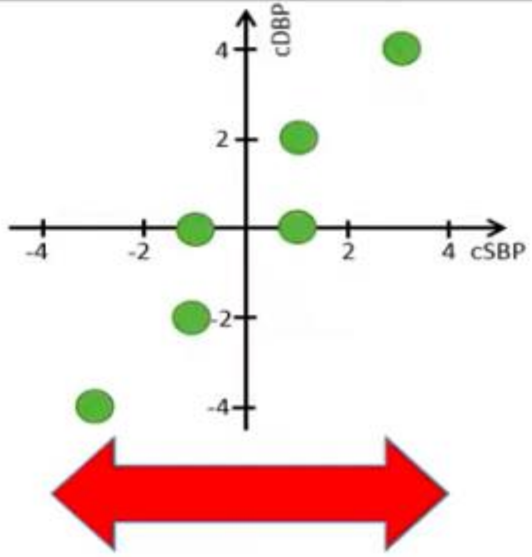


	SBP	DBP
SBP	4.4	5.6
DBP	5.6	$\boxed{8.0}$

We see that the spread in the diastolic blood pressure is a bit higher compared to

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4

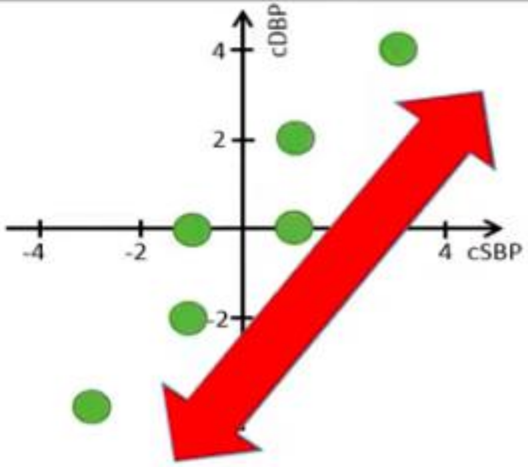


	SBP	DBP
SBP	$\boxed{4.4}$	5.6
DBP	5.6	8.0

the spread in the systolic blood pressure.

2. Calculate the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4

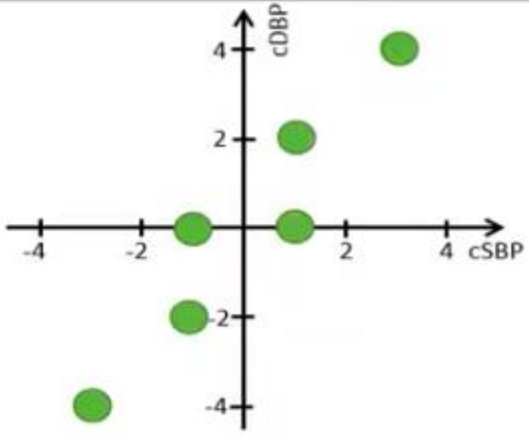


	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

The covariance is somewhere between these two values.

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



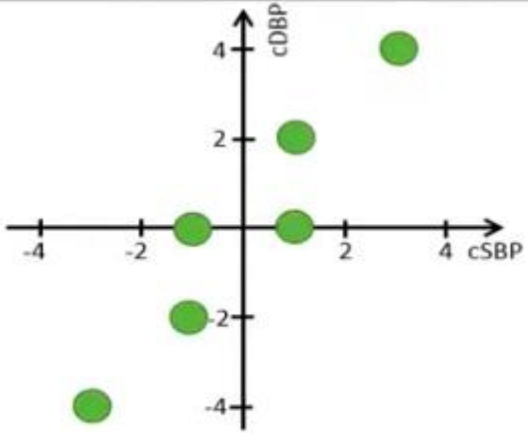
$$\boxed{\det|A - \lambda I| = 0}$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

Next, we calculate the eigenvalues of the covariance matrix.

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$\det\left|\boxed{A}-\lambda I\right|=0$$

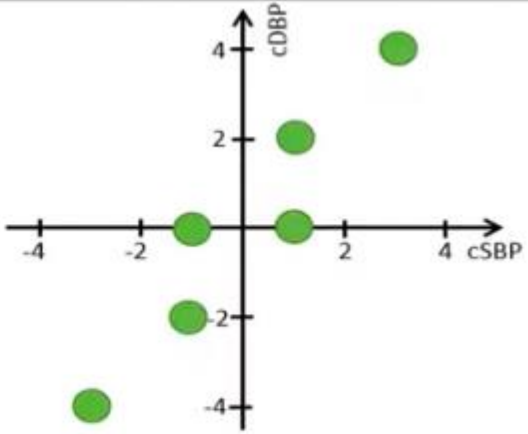
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\det\left|\boxed{\begin{bmatrix}4.4 & 5.6\\ 5.6 & 8.0\end{bmatrix}}-\begin{bmatrix}\lambda & 0\\ 0 & \lambda\end{bmatrix}\right|=0$$

We substitute A by the covariance matrix,

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$\det\left|A - \boxed{\lambda I}\right| = 0$$

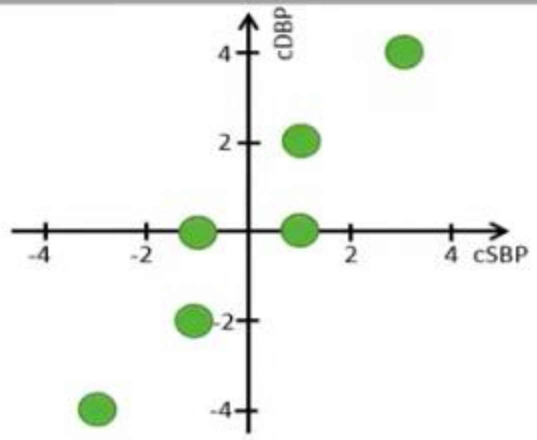
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\det\left|\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} - \boxed{\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}}\right| = 0$$

and this term by lambda times the identity matrix, which has the same number of rows and columns as the covariance matrix.

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$\det \boxed{A - \lambda I} = 0$$

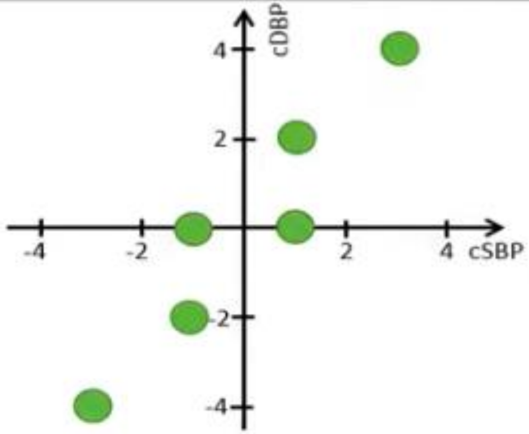
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\det \boxed{\left[\begin{bmatrix} (4.4-\lambda) & 5.6 \\ 5.6 & (8.0-\lambda) \end{bmatrix}\right]} = 0$$

Subtracting these two matrices results in the following matrix.

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$\boxed{\det}\left|A - \lambda I\right| = 0$$

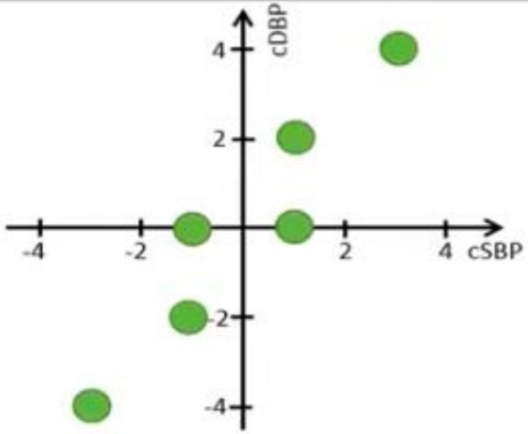
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\boxed{\det}\left|\begin{bmatrix} (4.4-\lambda) & 5.6 \\ 5.6 & (8.0-\lambda) \end{bmatrix}\right| = 0$$

Next, we calculate the determinant of this matrix,

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$\det|A - \lambda I| = 0$$

$$\boxed{(4.4-\lambda)(8.0-\lambda)} - 5.6 \cdot 5.6 = 0$$

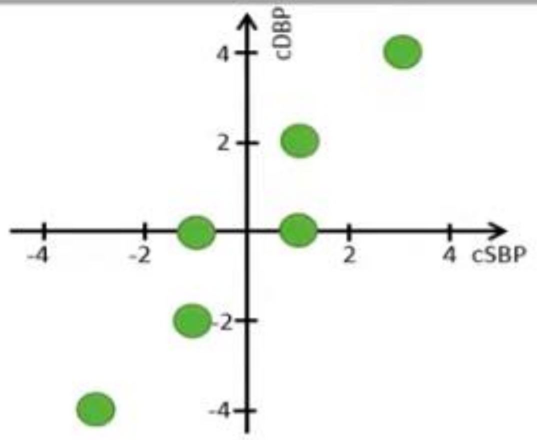
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\det\left[\begin{bmatrix} (4.4-\lambda) & 5.6 \\ 5.6 & (8.0-\lambda) \end{bmatrix}\right] = 0$$

which is the product of this diagonal,

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$\det\left|A-\lambda I\right|=0$$

$$(4.4-\lambda)(8.0-\lambda)\boxed{-}5.6\cdot 5.6=0$$

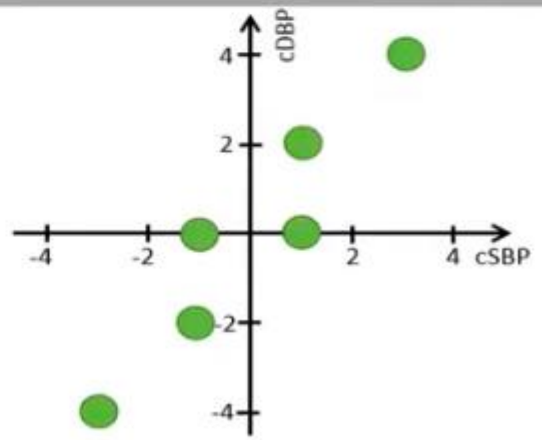
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\det\left[\begin{bmatrix}(4.4-\lambda) & 5.6 \\ 5.6 & (8.0-\lambda)\end{bmatrix}\right]=0$$

minus,

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$\det|A - \lambda I| = 0$$

$$(4.4-\lambda)(8.0-\lambda) - \boxed{5.6 \cdot 5.6} = 0$$

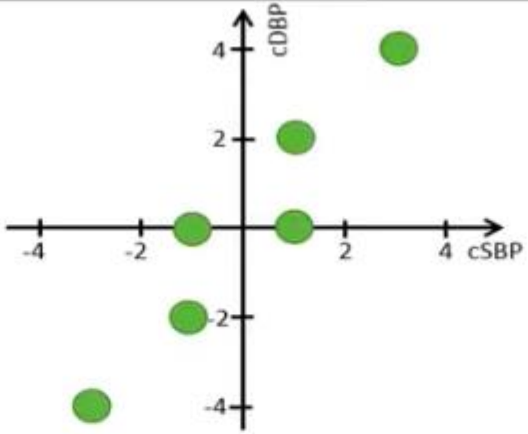
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\det\left|\begin{bmatrix} (4.4-\lambda) & 5.6 \\ 5.6 & (8.0-\lambda) \end{bmatrix}\right| = 0$$

the product of this diagonal.

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$\det|A - \lambda I| = 0$$

$$(4.4 - \lambda)(8.0 - \lambda) - 5.6 \cdot 5.6 = 0$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

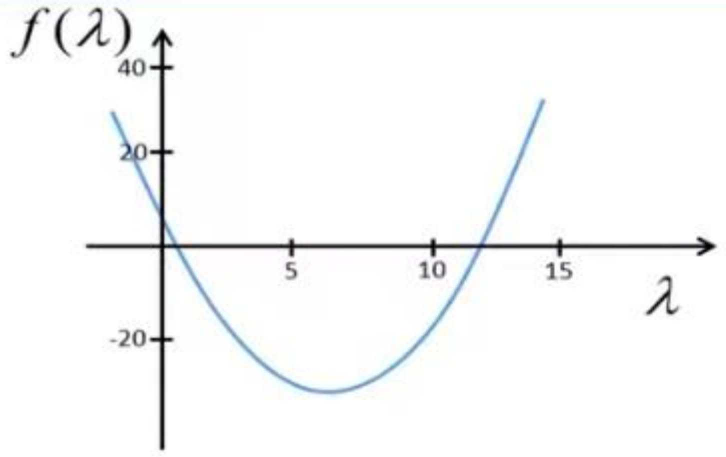
$$\det\left|\begin{bmatrix} (4.4-\lambda) & 5.6 \\ 5.6 & (8.0-\lambda) \end{bmatrix}\right| = 0$$

$$\boxed{3.84 - 12.4\lambda + \lambda^2 = 0}$$

After some simplifications, we have the following quadratic equation. Quadratic equations like this can be solved in different ways, which will not be discussed here.

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



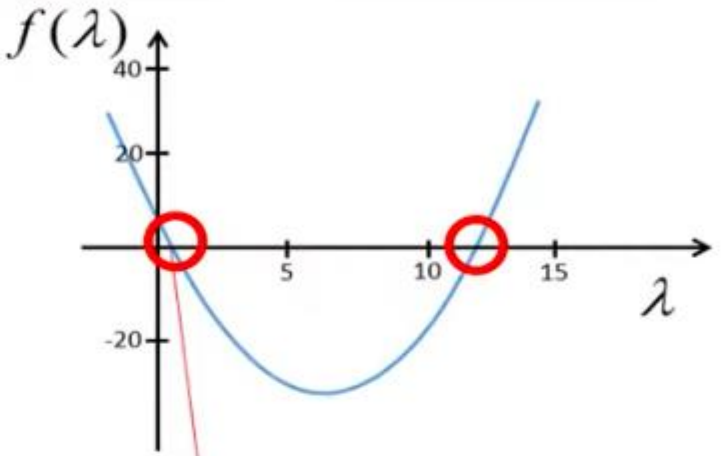
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\boxed{3.84 - 12.4\lambda + \lambda^2} = 0$$

However, if we plot how the left-hand side changes as a function of different values of lambda, we see that the left-hand side is equal to zero when lambda is equal to either,

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

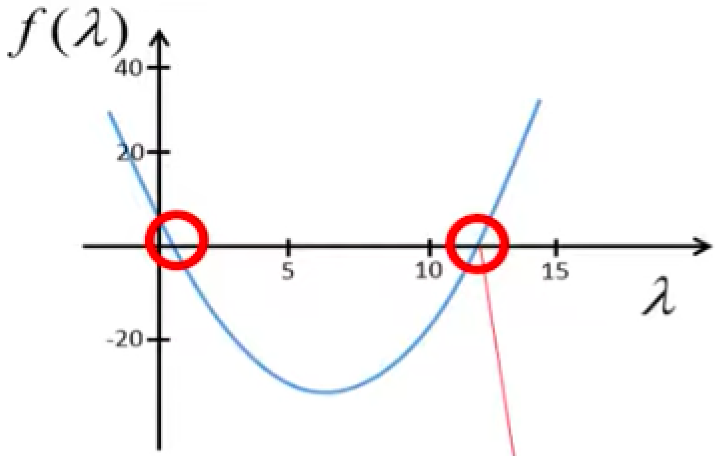
$$3.84 - 12.4\lambda + \lambda^2 = 0$$

$$\boxed{\lambda_1 = 0.32} \qquad \lambda_2 = 12.08$$

about 0.32,

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

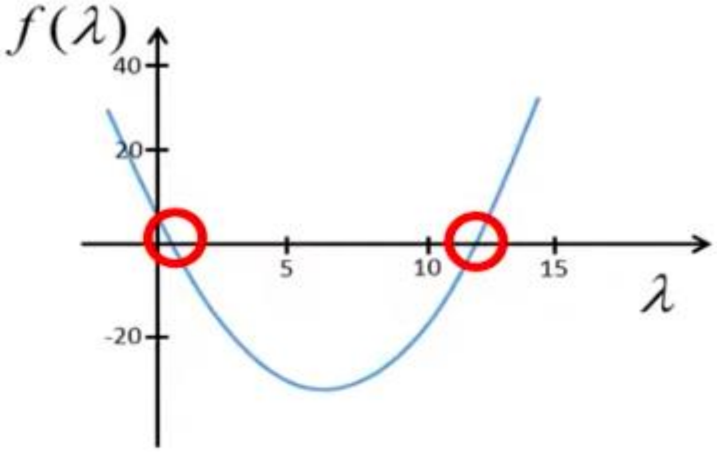
$$3.84 - 12.4\lambda + \lambda^2 = 0$$

$$\lambda_1 = 0.32 \qquad \boxed{\lambda_2 = 12.08}$$

or 12.08.

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

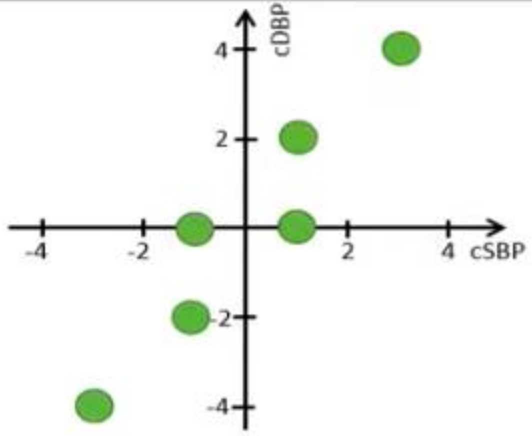
$$\boxed{3.84 - 12.4\lambda + \lambda^2} = 0$$

$$\lambda_1 = 0.32 \qquad \lambda_2 = 12.08$$

This means that if we set lambda to either 0.32 or 12.08, the left-hand side of this equation will become equal to zero, or close to zero due to rounding effects in this example.

3. Calculate the eigenvalues of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



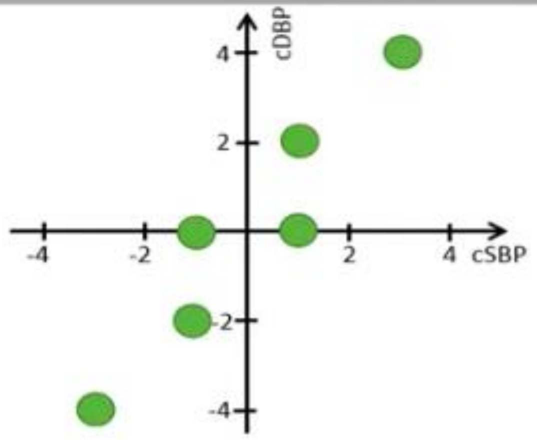
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\lambda_1 = 0.32 \qquad \lambda_2 = 12.08$$

These two values represent our eigenvalues of the covariance matrix.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$A \cdot v = \lambda \cdot v$$

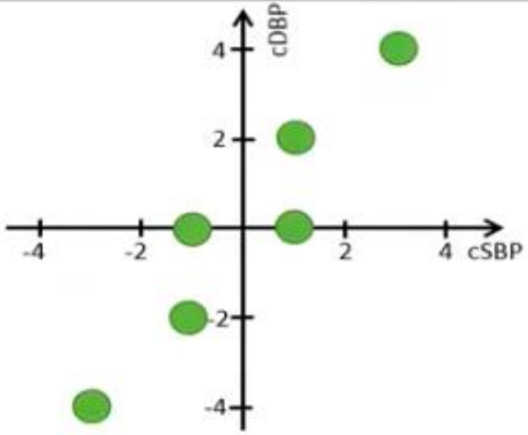
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$\lambda_1 = 0.32$ $\boxed{\lambda_2 = 12.08}$

Next, we calculate the corresponding eigenvectors to these two eigenvalues. We will start by calculating the eigenvector of the covariance matrix with the corresponding eigenvalue 12.08.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



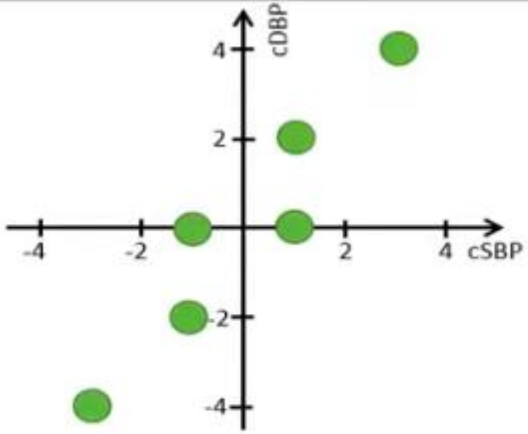
$$\boxed{A \cdot v = \lambda \cdot v}$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

To calculate the eigenvector of the covariance matrix, we use the following equation:

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$A \cdot v = \lambda \cdot v$$

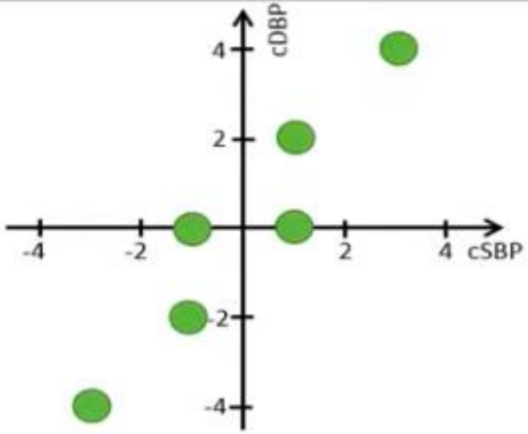
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\boxed{\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix}} \begin{bmatrix} x \\ y \end{bmatrix} = 12.08 \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

We plug in the covariance matrix,

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$A \cdot v = \lambda \cdot v$$

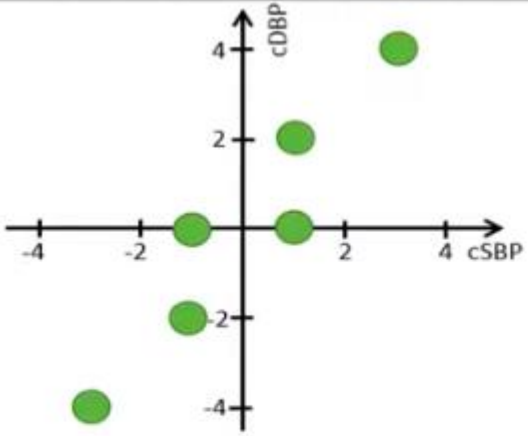
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \boxed{12.08} \begin{bmatrix} x \\ y \end{bmatrix}$$

and one of the two eigenvalues.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$A \cdot v = \lambda \cdot v$$

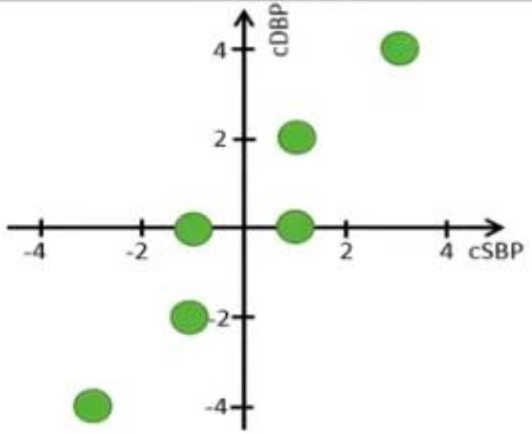
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \boxed{12.08 \cdot \begin{bmatrix} x \\ y \end{bmatrix}}$$

and multiply the eigenvalue by the same vector,

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$A \cdot v = \lambda \cdot v$$

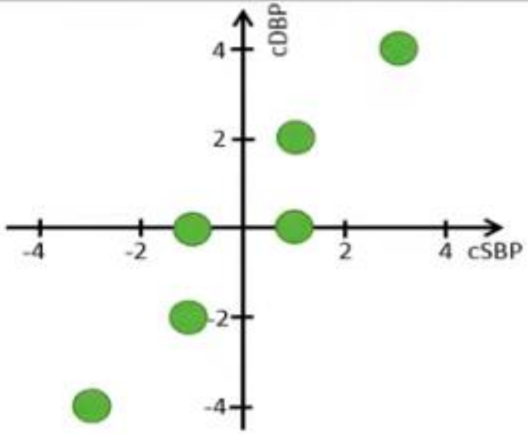
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \boxed{12.08 \cdot \begin{bmatrix} x \\ y \end{bmatrix}}$$

and multiply the eigenvalue by the same vector,

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$A \cdot v = \lambda \cdot v$$

$$4.4x + 5.6y = 12.08x$$
$$5.6x + 8.0y = 12.08y$$

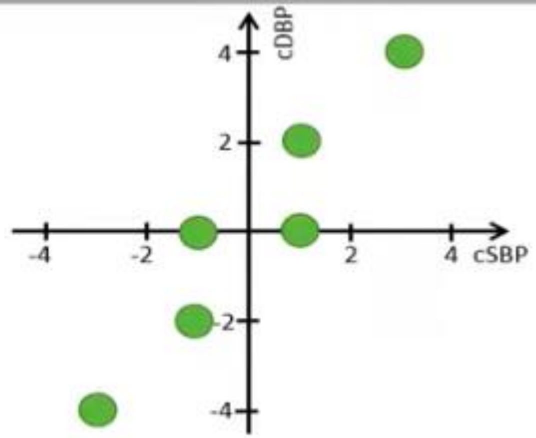
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 12.08 \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

we will get the following system of equations.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$A \cdot v = \lambda \cdot v$$

$$\boxed{4.4x} + 5.6y = 12.08x$$
$$5.6x + \boxed{8.0y} = 12.08y$$

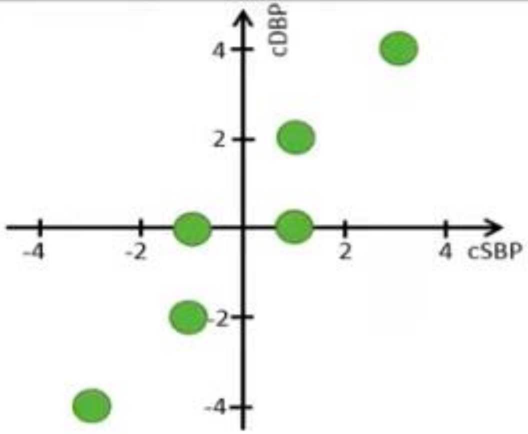
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 12.08 \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

We move these two terms to the right-hand side.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$4.4x + 5.6y = 12.08x$$
$$5.6x + 8.0y = 12.08y$$

$$A \cdot v = \lambda \cdot v$$

$$\boxed{\begin{aligned} 5.6y &= 7.68x \\ 5.6x &= 4.08y \end{aligned}}$$

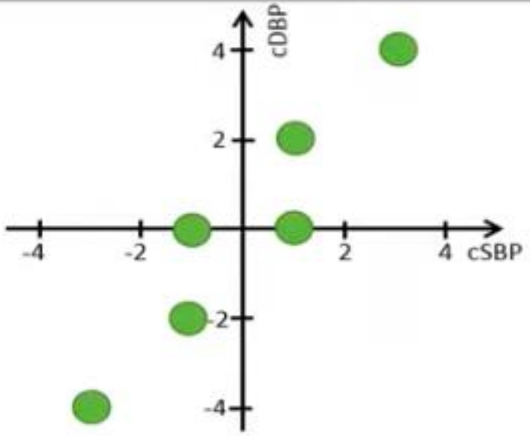
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 12.08 \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

After some simplifications, we have the following system of equations.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$A \cdot v = \lambda \cdot v$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 12.08 \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

$$4.4x + 5.6y = 12.08x$$

$$5.6x + 8.0y = 12.08y$$

$$5.6y = 7.68x$$

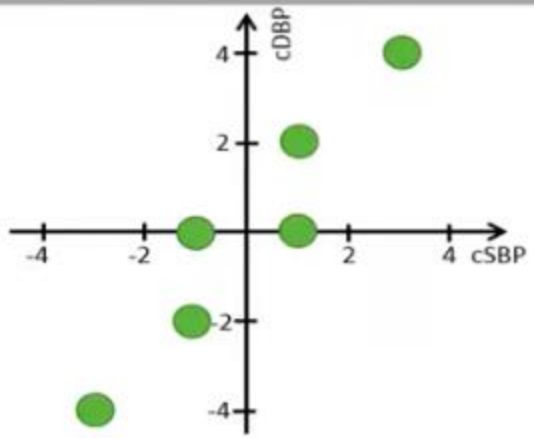
$$5.6x = 4.08y$$

$$\boxed{\begin{aligned} y &= 1.37x \\ 1.37x &= y \end{aligned}}$$

Solving for y in the two equations, results in that y is equal to 1.37 x.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$A \cdot v = \lambda \cdot v$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 12.08 \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

$$4.4x + 5.6y = 12.08x$$

$$5.6x + 8.0y = 12.08y$$

$$5.6y = 7.68x$$

$$5.6x = 4.08y$$

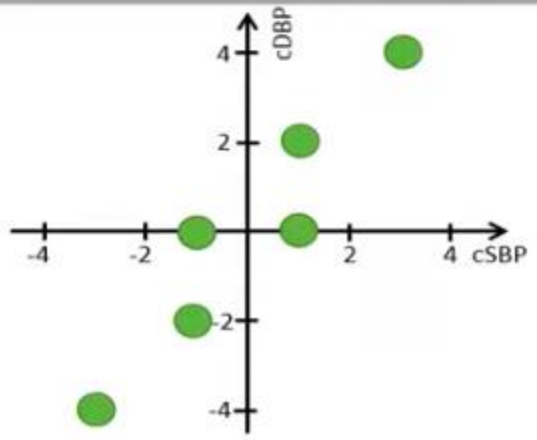
$$y = 1.37\boxed{x}$$

$$1.37x = y$$

For example, if we set x equal to one,

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$A \cdot v = \lambda \cdot v$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 12.08 \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

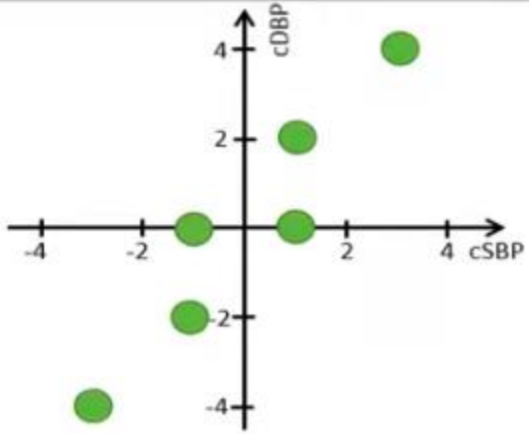
$$4.4x + 5.6y = 12.08x$$
$$5.6x + 8.0y = 12.08y$$

$$5.6y = 7.68x$$
$$5.6x = 4.08y$$
$$\boxed{y} = 1.37x$$
$$1.37x = y$$

y is equal to 1.37.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



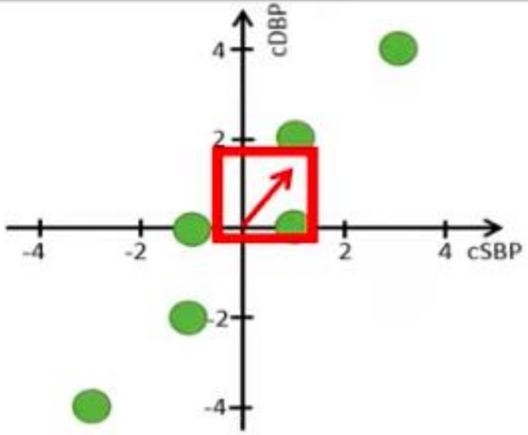
$$\boxed{v_2 = \begin{bmatrix} 1 \\ 1.37 \end{bmatrix}}$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

This vector is therefore an eigenvector of the covariance matrix.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



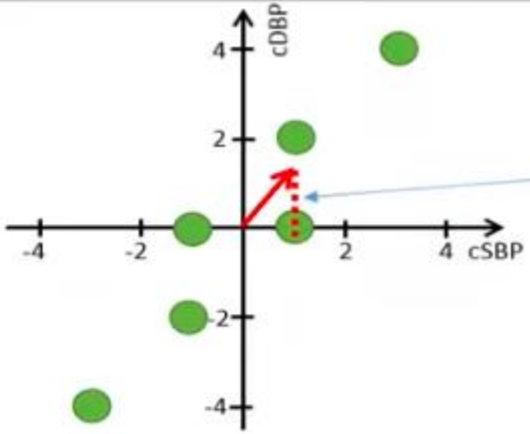
$$v_2 = \begin{bmatrix} 1 \\ 1.37 \end{bmatrix}$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

We can illustrate this vector in the plot like this, by drawing an arrow from the origin to the coordinates,

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



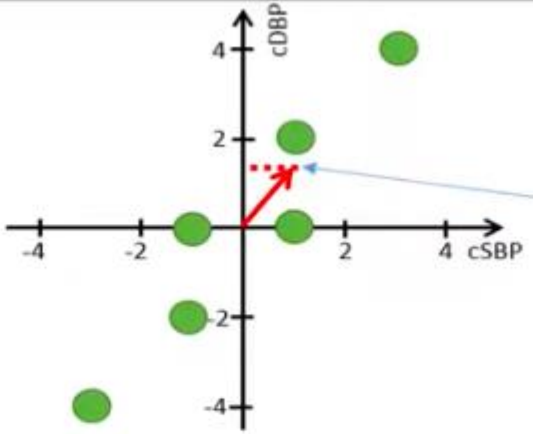
$$v_2 = \begin{bmatrix} \boxed{1} \\ 1.37 \end{bmatrix}$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

one

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



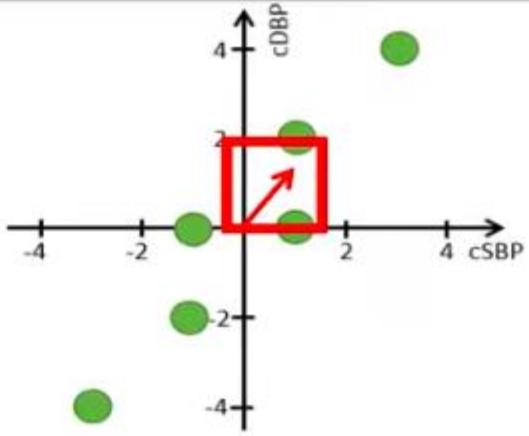
$$v_2 = \begin{bmatrix} 1 \\ \boxed{1.37} \end{bmatrix}$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

and 1.37.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



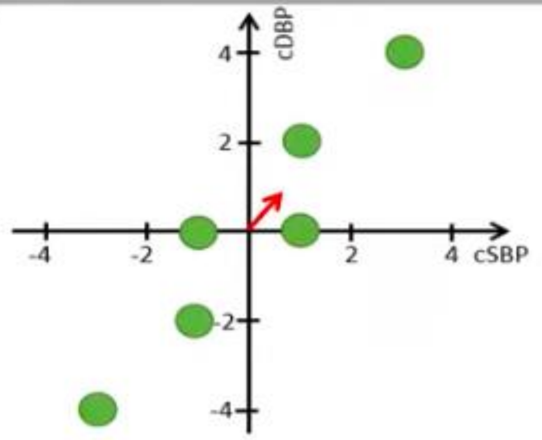
$$v_2 = \begin{bmatrix} 1 \\ 1.37 \end{bmatrix}$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

We will now normalize this vector to unit length, which means that it should have a length of one.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



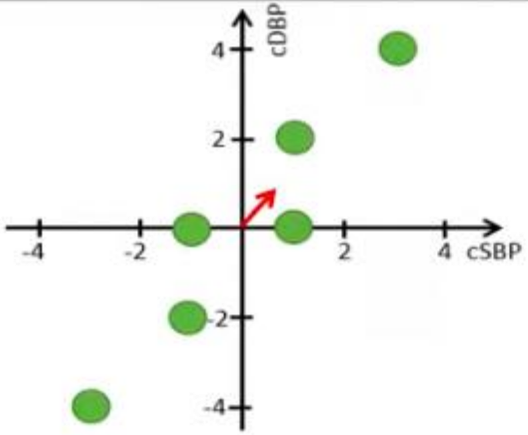
$$\boxed{v_2 = \begin{bmatrix} 0.59 \\ 0.81 \end{bmatrix}} \quad \lambda_2 = 12.08$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

After normalization, this vector represents one out of two eigenvectors of the covariance matrix.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$v_2 = \begin{bmatrix} 0.59 \\ 0.81 \end{bmatrix} \quad \lambda_2 = 12.08$$

$$\lambda_1 = 0.32$$

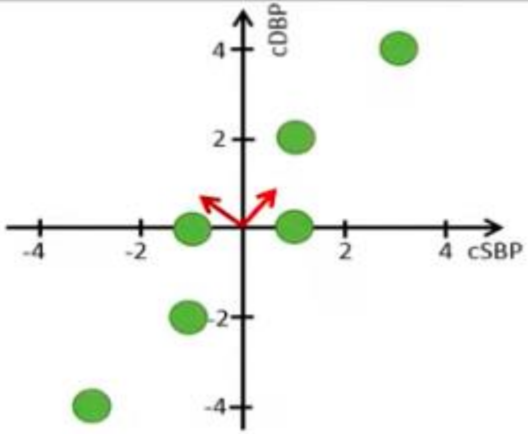
	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \boxed{0.32} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

To find the second eigenvector, we do the same calculations as before based on the second eigenvalue.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$v_2 = \begin{bmatrix} 0.59 \\ 0.81 \end{bmatrix} \quad \lambda_2 = 12.08$$

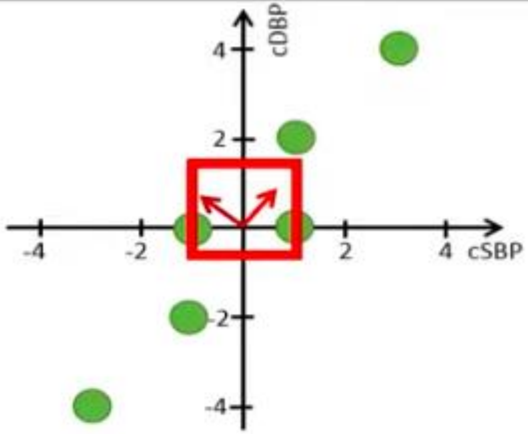
$$\boxed{v_1 = \begin{bmatrix} -0.81 \\ 0.59 \end{bmatrix}} \quad \lambda_1 = 0.32$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

After some calculations, this vector represents our second eigenvector with unit length.

4. Calculate the eigenvectors of the covariance matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$v_2 = \begin{bmatrix} 0.59 \\ 0.81 \end{bmatrix} \quad \lambda_2 = 12.08$$

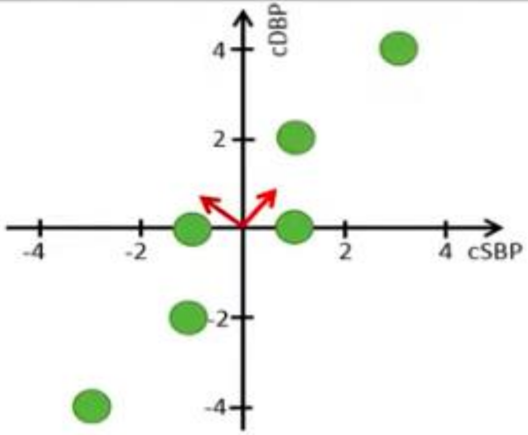
$$v_1 = \begin{bmatrix} -0.81 \\ 0.59 \end{bmatrix} \quad \lambda_1 = 0.32$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

Since the covariance matrix is a symmetric matrix, the eigenvectors will be orthogonal, which means that the angle between them is 90 degrees.

5. Order the eigenvectors

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$v_2 = \begin{bmatrix} 0.59 \\ 0.81 \end{bmatrix} \qquad \lambda_2 = 12.08$$

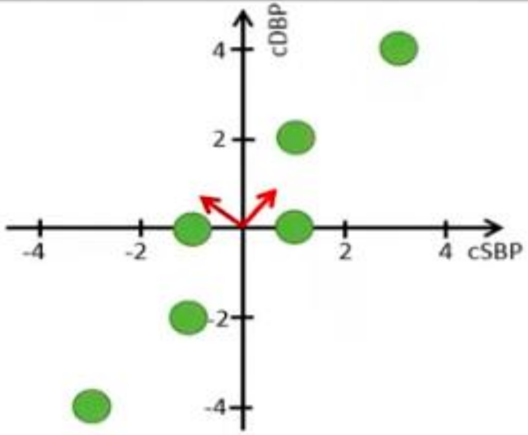
$$v_1 = \begin{bmatrix} -0.81 \\ 0.59 \end{bmatrix} \qquad \lambda_1 = 0.32$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

Next, we order the eigenvectors based on their corresponding eigenvalues, where the eigenvector with the largest eigenvalue becomes our first eigenvector.

5. Order the eigenvectors

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$\boxed{v_1 = \begin{bmatrix} 0.59 \\ 0.81 \end{bmatrix} \quad \lambda_1 = 12.08}$$

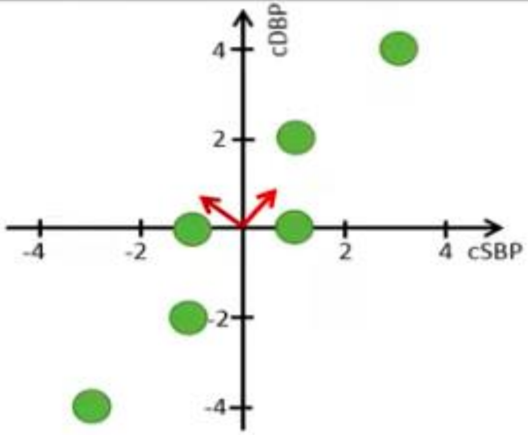
$$v_2 = \begin{bmatrix} -0.81 \\ 0.59 \end{bmatrix} \quad \lambda_2 = 0.32$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

Since this eigenvector has the largest eigenvalue, it will represent our first eigenvector. We therefore rename this vector so that it is called v1 instead of v2.

5. Order the eigenvectors

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$v_1 = \begin{bmatrix} 0.59 \\ 0.81 \end{bmatrix} \quad \lambda_1 = 12.08$$

$$v_2 = \begin{bmatrix} -0.81 \\ 0.59 \end{bmatrix} \quad \lambda_2 = 0.32$$

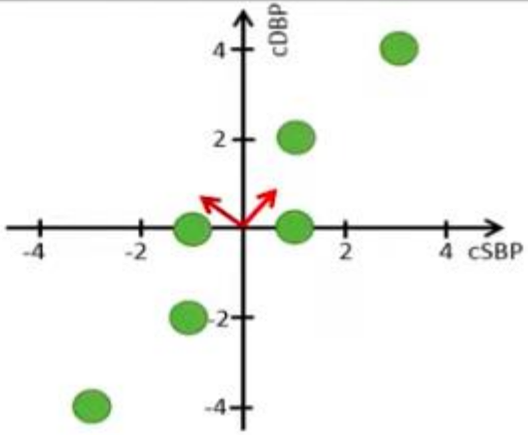
$$\boxed{V = \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix}}$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

Let's put these two eigenvectors together into a matrix that we call V,

5. Order the eigenvectors

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$v_1 = \begin{bmatrix} 0.59 \\ 0.81 \end{bmatrix} \quad \lambda_1 = 12.08$$

$$v_2 = \begin{bmatrix} -0.81 \\ 0.59 \end{bmatrix} \quad \lambda_2 = 0.32$$

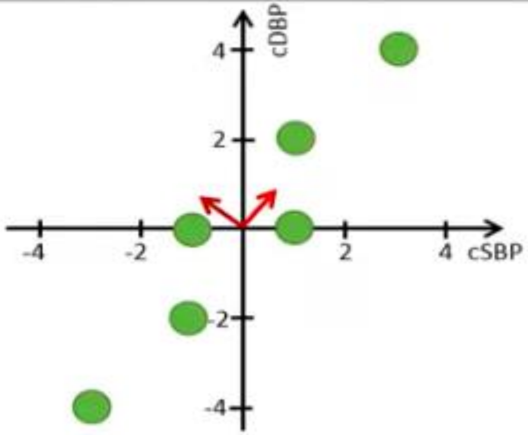
$$V = \begin{bmatrix} \boxed{\begin{matrix} 0.59 \\ 0.81 \end{matrix}} & \begin{matrix} -0.81 \\ 0.59 \end{matrix} \end{bmatrix}$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

where the first column represents the first eigenvector with the highest eigenvalue,

5. Order the eigenvectors

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



$$v_1 = \begin{bmatrix} 0.59 \\ 0.81 \end{bmatrix} \quad \lambda_1 = 12.08$$

$$v_2 = \begin{bmatrix} -0.81 \\ 0.59 \end{bmatrix} \quad \lambda_2 = 0.32$$

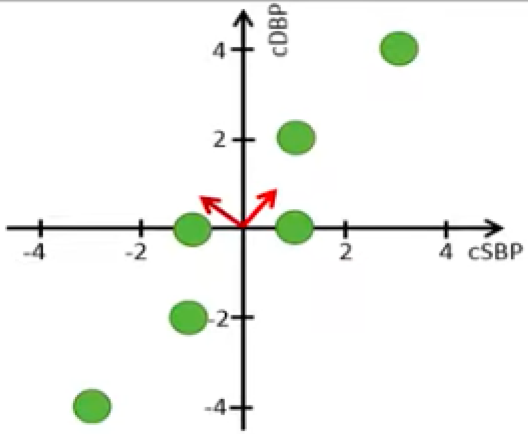
$$V = \begin{bmatrix} 0.59 & \boxed{\begin{matrix} -0.81 \\ 0.59 \end{matrix}} \\ 0.81 & \end{bmatrix}$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

and the second column represents our second eigenvector.

6. Calculate the principal components

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4

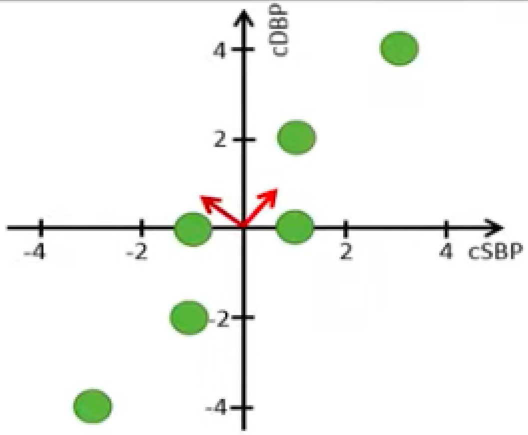


$$V = \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix}$$

We will now use this matrix to transform our original centered data so that the two variables are completely uncorrelated.

6. Calculate the principal components

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



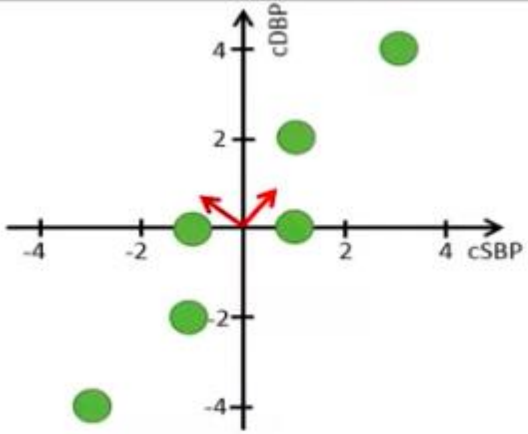
$$V = \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix}$$

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

We define a matrix D, which includes our centered data.

6. Calculate the principal components

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



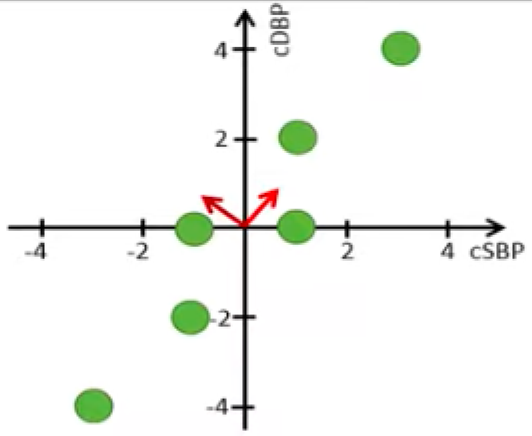
$$V = \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix} \qquad D = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix}$$

$$\boxed{DV = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix}} = \begin{bmatrix} -5.0 & 0.1 \\ -2.2 & -0.4 \\ -0.6 & 0.8 \\ 0.6 & -0.8 \\ 2.2 & 0.4 \\ 5.0 & -0.1 \end{bmatrix}$$

Next, we multiply our data matrix D by matrix V, which includes our eigenvectors as columns.

6. Calculate the principal components

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4

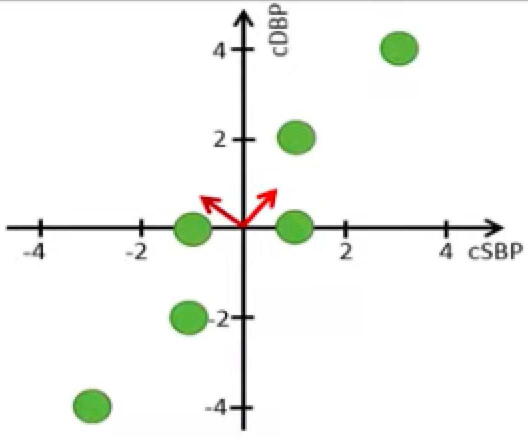


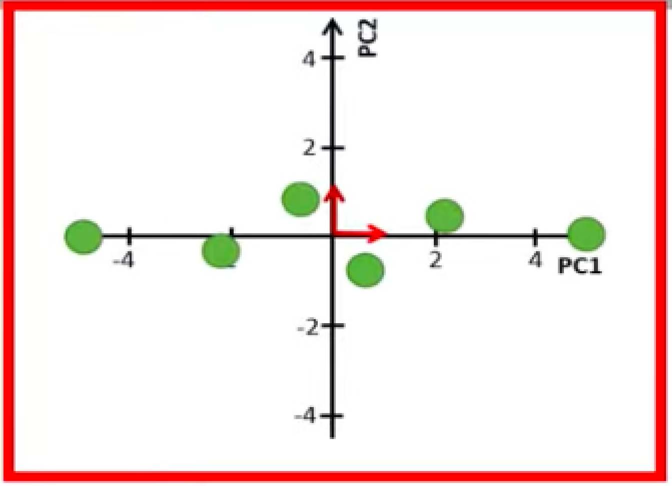
$$V = \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix} \qquad D = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad DV = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix} = \boxed{\begin{bmatrix} -5.0 & 0.1 \\ -2.2 & -0.4 \\ -0.6 & 0.8 \\ 0.6 & -0.8 \\ 2.2 & 0.4 \\ 5.0 & -0.1 \end{bmatrix}}$$

Then we get a new matrix with the transformed data. This transformed data is called principal component scores, or just scores, which represent the original centered data in the principal component space.

6. Calculate the principal components

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4





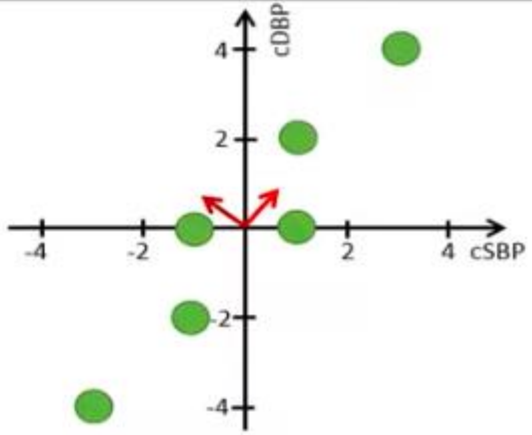
$$V = \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix} \quad D = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix}$$

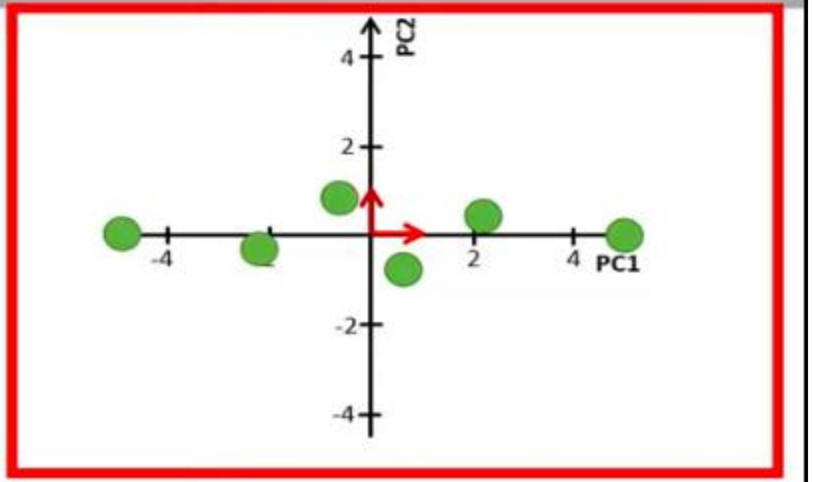
$$DV = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix} = \begin{bmatrix} -5.0 & 0.1 \\ -2.2 & -0.4 \\ -0.6 & 0.8 \\ 0.6 & -0.8 \\ 2.2 & 0.4 \\ 5.0 & -0.1 \end{bmatrix}$$

When we go from our original data matrix to the transformed data, this can be seen like we rotate the data clockwise until the two eigenvectors point in the same direction as the x and y-axes of the plot.

6. Calculate the principal components

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4





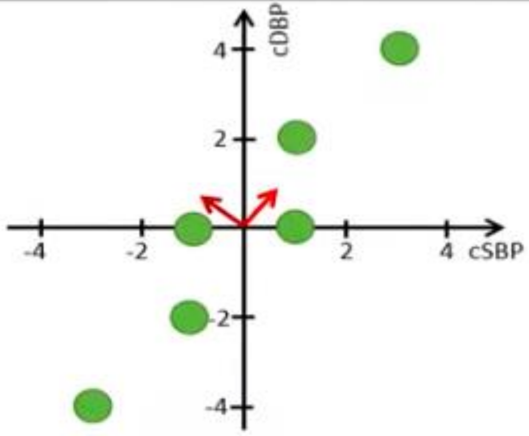
$$V = \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix} \qquad D = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix}$$

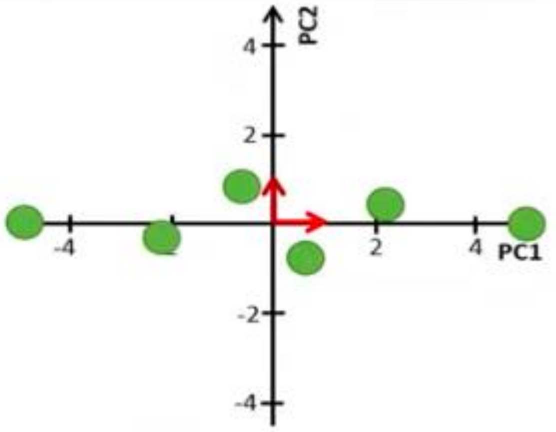
$$DV = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix} = \begin{bmatrix} -5.0 & 0.1 \\ -2.2 & -0.4 \\ -0.6 & 0.8 \\ 0.6 & -0.8 \\ 2.2 & 0.4 \\ 5.0 & -0.1 \end{bmatrix}$$

The rotated data now looks like this. Note that the labels of the axes have now been changed to principal component one and two.

6. Calculate the principal components

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4





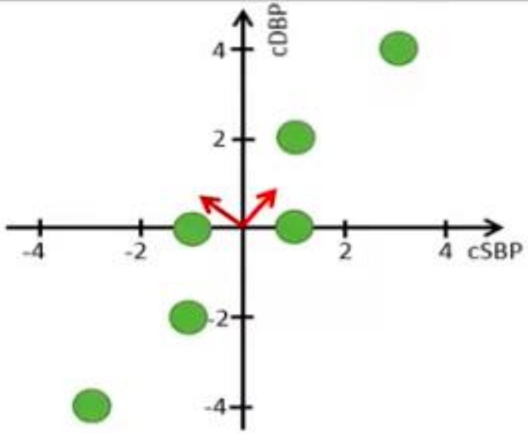
$$V = \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix} \qquad D = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix}$$

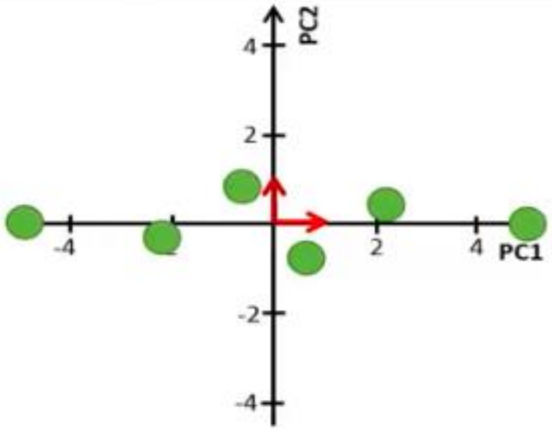
$$DV = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix} = \begin{array}{c} \begin{array}{cc} \textbf{PC1} & \textbf{PC2} \end{array} \\ \boxed{\begin{bmatrix} -5.0 & 0.1 \\ -2.2 & -0.4 \\ -0.6 & 0.8 \\ 0.6 & -0.8 \\ 2.2 & 0.4 \\ 5.0 & -0.1 \end{bmatrix}} \end{array}$$

Let's call the two columns of the transformed data PC1 and PC2.

6. Calculate the principal components

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4





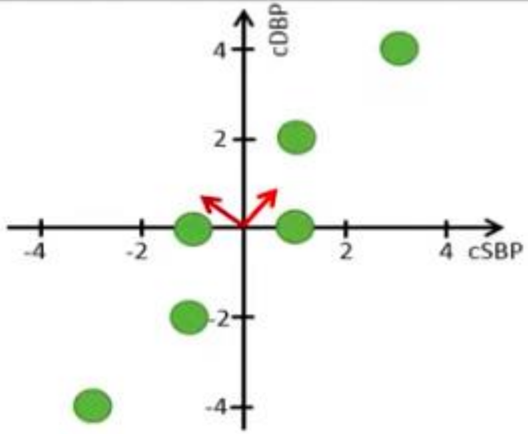
$$V = \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix} \qquad D = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix}$$

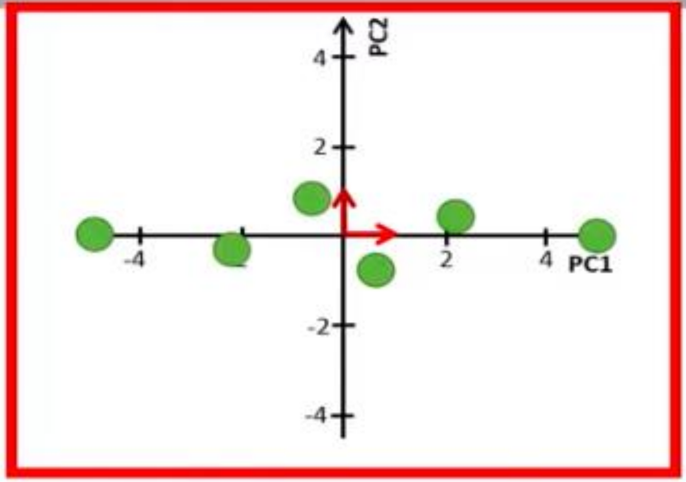
$$DV = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix} = \boxed{\begin{array}{c} \begin{array}{cc} \textbf{PC1} & \textbf{PC2} \end{array} \\ \begin{bmatrix} -5.0 & 0.1 \\ -2.2 & -0.4 \\ -0.6 & 0.8 \\ 0.6 & -0.8 \\ 2.2 & 0.4 \\ 5.0 & -0.1 \end{bmatrix} \end{array}}$$

If we would plot this data, where we label the x-axis as PC1 and the y-axis as PC2,

6. Calculate the principal components

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4





$$V=\begin{bmatrix}0.59 & -0.81\\ 0.81 & 0.59\end{bmatrix} \quad D=\begin{bmatrix}-3 & -4\\ -1 & -2\\ -1 & 0\\ 1 & 0\\ 1 & 2\\ 3 & 4\end{bmatrix}$$

$$DV=\begin{bmatrix}-3 & -4\\ -1 & -2\\ -1 & 0\\ 1 & 0\\ 1 & 2\\ 3 & 4\end{bmatrix}\cdot\begin{bmatrix}0.59 & -0.81\\ 0.81 & 0.59\end{bmatrix}=\begin{array}{cc} \text{PC1} & \text{PC2} \\ \end{array}\begin{bmatrix}-5.0 & 0.1\\ -2.2 & -0.4\\ -0.6 & 0.8\\ 0.6 & -0.8\\ 2.2 & 0.4\\ 5.0 & -0.1\end{bmatrix}$$

we would get the following plot, which represents the original plot after the rotation. Since we plot the principal component scores, this kind of plot is called a score plot.